TESIS DE LA UNIVERSIDAD DE ZARAGOZA
2025
216

Inés Medina Lozano

Nutritional value and response to drought stress mediated by metabolites of lettuce (Lactuca sativa L.) and wild relatives using biochemical, genetic, genomic and transcriptomic tools

Director/es
Díaz Bermúdez, Aurora

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

NUTRITIONAL VALUE AND RESPONSE TO DROUGHT STRESS MEDIATED BY METABOLITES OF LETTUCE (LACTUCA SATIVA L.) AND WILD RELATIVES USING BIOCHEMICAL, GENETIC, GENOMIC AND TRANSCRIPTOMIC TOOLS

Autor

Inés Medina Lozano

Director/es

Díaz Bermúdez, Aurora

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Ciencias Agrarias y del Medio Natural

2025

Universidad Zaragoza
1542

Tesis Doctoral

Nutritional value and response to drought stress mediated by metabolites of lettuce (*Lactuca sativa* L.) and wild relatives using biochemical, genetic, genomic and transcriptomic tools

Autor

Inés Medina Lozano

Director/es

Aurora Díaz Bermúdez

Escuela de Doctorado
2024

Nutritional value and response to drought stress mediated by metabolites of lettuce (*Lactuca sativa* L.) and wild relatives using biochemical, genetic, genomic and transcriptomic tools

Inés Medina Lozano

DOCTORAL THESIS | 2024

DOCTORAL THESIS

Nutritional value and response to drought stress mediated by metabolites of lettuce (*Lactuca sativa* L.) and wild relatives using biochemical, genetic, genomic and transcriptomic tools

Presented by

Inés Medina Lozano

To opt for the Degree of Doctor from the University of Zaragoza

Doctoral Programme of Agricultural and Environmental Sciences

Director:

Aurora Díaz Bermúdez

Zaragoza, December 2024

Zaragoza, 18th of December 2024

Aurora Díaz Bermúdez, PhD in Genetic Breeding and Engineering from the University of Córdoba, reports that the work titled "Nutritional value and response to drought stress mediated by metabolites of lettuce (*Lactuca sativa* L.) and wild relatives using biochemical, genetic, genomic and transcriptomic tools", carried out by Inés Medina Lozano under her direction, has sufficient merits to qualify for the Degree of Doctor in Agricultural and Environmental Sciences from the University of Zaragoza.

Considering that it is completed, she gives the approval for its presentation and defence as a Doctoral Thesis in the Department of Agricultural and Environmental Sciences of the University of Zaragoza.

Signed: Aurora Díaz Bermúdez

Agricultural Researcher

[1]Department of Plant Science, Agrifood Research and Technology Centre of Aragon (CITA)

Avda. Montañana 930, 50059 Zaragoza (Spain)

[2]AgriFood Institute of Aragon – IA2 (CITA-University of Zaragoza)

50013 Zaragoza (Spain)

COMPENDIUM OF PUBLICATIONS

This Doctoral Thesis is presented as a compendium of five research papers previously published. The publications that constitute the main body of the Thesis are detailed below:

1. Medina-Lozano, I, Bertolín, JR, Díaz, A (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* 359, 129864. https://doi.org/10.1016/j.foodchem.2021.129864

2. Medina-Lozano, I, Bertolín, JR, Plieske, J, Ganal, M, Gnad, H, Díaz, A (2024b). Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays. *Plant Genome* 17, e20518. https://doi.org/10.1002/tpg2.20518

3. Medina-Lozano, I, Bertolín, JR, Díaz, A (2024a). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15, 3389. https://doi.org/10.3389/fpls.2024.1369658

4. Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2023). Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives. *Int. J. Mol. Sci.* 24, 3052. https://doi.org/10.3390/ijms24033052

5. Medina-Lozano, I, Grimplet, J, Díaz, A (2025). Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress. *Front. Plant Sci.* 15, 1494339. https://doi.org/10.3389/fpls.2024.1494339

FUNDING

This work was funded by the projects RTA2017-00093-00-00 from the National Institute for Agricultural and Food Research and Technology; LMP164_18 and LMP148_21 from the Government of Aragón; PID2022-138484OR-I00 from the Spanish Ministry of Science and Innovation (MCIN) and the State Research Agency (AEI); AGROALNEXT from the MCIN with funding of the European Union NextGenerationEU: PRTR-C17.11; and by the Operational Programme FEDER Aragón 2014-2020, 2020-2022 and 2023-2025, and the European Social Fund from the European Union (A12-17R, A12_20R and A12_23R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética").

Inés Medina Lozano was supported by the predoctoral contract for training doctors PRE2018-084987 from the Spanish Ministry of Science, Innovation and Universities and the Spanish State Research Agency.

The seeds used in this thesis have been supplied by:

- Vegetable Germplasm Bank of Zaragoza (BGHZ-CITA, Spain).
- Centre for Genetic Resources (CGN, Wageningen, Netherlands).
- Ramiro Arnedo Semillas S.A.

INDEX

ABSTRACT

Lettuce (*Lactuca sativa* L.), one of the most popular leafy vegetables worldwide, is widely considered a healthy food. In recent years, its production has increased likely due to the growing interest of consumers in the impact of diet on health. However, lettuce breeding programmes have mainly been focused on disease resistance and yield increase, putting aside the enhancement of health beneficial properties or nutritional quality. Tolerance to abiotic stresses is another neglected aspect in lettuce breeding programmes, despite its importance in the current climate change scenario. Therefore, the objectives of this thesis are, on one hand, to study the genetic basis of vitamin C content, the main indicator of nutritional value, as the beginning of a future biofortification project and, on the other hand, to study the drought stress impact on vitamin C and anthocyanin content, two powerful antioxidant compounds, in lettuce germplasm and wild relative species. To achieve these goals, the content of these compounds was firstly quantified in commercial (10) and traditional (13) varieties, as well as in wild relatives (7). The highest average content in vitamin C was found in the wild species and the lowest in the commercial varieties. On the contrary, the commercial varieties showed the highest average concentration of anthocyanins and the wild relatives, the lowest. The traditional varieties showed an intermediate content of both compounds. Nevertheless, considering raw data, the highest content in vitamin C was found in one plant of the traditional variety 'Lechuga del Pirineo'. These results served as the starting point for the two research lines pursued in the current thesis.

On one hand, the accession 'Lechuga del Pirineo' was selected with the aim of eventually obtaining a uniform variety rich in vitamin C. Genetic homogeneity was already reached in S1 population and ascorbic acid, the vitamin C form with the highest biological activity, increased significantly in S2 population (S1 and S2 were obtained by self-pollinating the richest plant in vitamin C from the original S0 population and S1, respectively). A genome-wide association study of vitamin C content was conducted in 205 plants of 'Lechuga del Pirineo' S0 and a diversity panel of 21 cultivated varieties. Significant genetic associations were found between 17 single nucleotide polymorphisms (SNPs) and the content in dehydroascorbic acid (the other form of vitamin C) in a 5.1 Mb region of chromosome 2. Among them, 12 showed high linkage disequilibrium with the most significantly associated SNP. Genes putatively related to vitamin C content were identified in that region, including some that have previously been described to participate in the regulation of this compound in other crops, such as tomato, potato and wheat.

On the other hand, a set of five cultivated lettuce varieties and two wild relatives were selected to evaluate the effect of drought stress on both vitamin C and anthocyanin

content for two years. Water deficit caused an increase in anthocyanin content in all the assayed accessions, unlike what happened with vitamin C amount, which decreased in all cases. This suggests that anthocyanins might be implied in drought tolerance in some *Lactuca* spp. and, for this reason, transcriptomic analyses were focused on the identification of differentially expressed genes (DEGs) specifically related to anthocyanins under water stress conditions. A commercial lettuce variety ('Romired') and a wild species (*Lactuca homblei* De Wild) were selected for the transcriptomic study (RNA-seq). *L. homblei* was the accession that showed both the highest enrichment in anthocyanin content (and the only in which it was statistically significant) among all the tested ones and the highest transcriptomic response when subject to stress. The number of upregulated DEGs, including those related to anthocyanin content, was higher in *L. homblei*, whereas the number of downregulated DEGs was higher in 'Romired'. Among DEGs showing the highest change in the expression level and a significant correlation with both anthocyanin content and drought stress treatment, 19 candidate genes were selected. Interestingly, these genes resulted to be differentially expressed exclusively in *L. homblei*. In addition, all of them have shown functions related to anthocyanin content and/or response to biotic and abiotic stresses in other plants. In the wild species sequences, some candidate genes had polymorphisms with high or moderate predicted impacts on their protein functions, like a heat shock protein that participates in the abscisic acid-induced stomatal closure, a transcription factor involved in responses to biotic and abiotic stresses and a phospholipase implied in anthocyanin accumulation under abiotic stress. To validate their expression by quantitative PCR (qPCR), the most stable reference gene for the drought stress experiment was previously selected based on the RNA-seq data. Expression patterns from both techniques (RNA-seq and qPCR) were consistent and, thus, the reliability of the transcriptomic study was confirmed. All results considered, *L. homblei* could be a valuable resource in lettuce breeding programmes as a source of drought stress tolerance and health beneficial compounds, making also possible a more sustainable lettuce cultivation.

RESUMEN

La lechuga (*Lactuca sativa* L.), una de las hortalizas de hoja más populares a nivel mundial, es ampliamente reconocida como un alimento saludable. En los últimos años, su producción ha aumentado debido probablemente al creciente interés de los consumidores por el impacto de la dieta en la salud. No obstante, los programas de mejora de lechuga se han venido centrando principalmente en la resistencia a enfermedades y en el incremento del rendimiento, dejando de lado la mejora de las propiedades beneficiosas para la salud o de la calidad nutricional. La tolerancia a estreses abióticos es otro de los aspectos olvidados en los programas de mejora de lechuga, a pesar de su importancia en el escenario actual de cambio climático. Por tanto, los objetivos de esta tesis son, por un lado, el estudio de la base genética del contenido en vitamina C, principal indicador del valor nutricional, como punto de partida para un proyecto futuro de biofortificación y, por otro lado, el estudio del impacto del estrés por sequía en el contenido en vitamina C y antocianinas, dos potentes compuestos antioxidantes, en germoplasma de lechuga y especies silvestres relacionadas. Para alcanzar estos objetivos, se cuantificó en primer lugar el contenido de estos compuestos en variedades comerciales (10) y tradicionales (13) de lechuga, así como en parientes silvestres (7). El contenido medio más alto en vitamina C se encontró en las especies silvestres y el más bajo en las variedades comerciales. Por el contrario, las variedades comerciales fueron las que mostraron la mayor concentración media de antocianinas y los parientes silvestres, la menor. Las variedades tradicionales mostraron un contenido intermedio de los dos compuestos. No obstante, teniendo en cuenta los datos brutos, el contenido más alto en vitamina C se encontró en una planta de la variedad tradicional 'Lechuga del Pirineo'. Estos resultados sentaron las bases para las dos líneas de investigación abordadas en esta tesis.

Por un lado, se seleccionó la accesión 'Lechuga del Pirineo' con el objetivo de obtener en el futuro una variedad uniforme rica en vitamina C. La homogeneidad genética se alcanzó en la población S1 y el contenido en ácido ascórbico, la forma de vitamina C con mayor actividad biológica, aumentó significativamente en la población S2 (S1 y S2 se obtuvieron por autofecundación de la planta más rica en vitamina C de la población original S0 y de S1, respectivamente). Se llevó a cabo un estudio de asociación del genoma completo del contenido en vitamina C en 205 plantas de la población S0 de 'Lechuga del Pirineo' y un panel de diversidad de 21 variedades cultivadas. Se encontraron asociaciones genéticas significativas entre 17 SNPs (single nucleotide polymorphisms) y el contenido en ácido dehidroascórbico (la otra forma de vitamina C) en una región de 5,1 Mb del cromosoma 2. Entre ellos, 12 mostraron un desequilibrio de ligamiento alto con el SNP más significativamente asociado. Se identificaron genes putativamente relacionados con el contenido en vitamina C en esa región, incluyendo

algunos cuya participación en la regulación de este compuesto se ha descrito previamente en otros cultivos, como tomate, patata y trigo.

Por otro lado, se seleccionaron cinco variedades de lechuga cultivada y dos parientes silvestres para evaluar el efecto del estrés por sequía en el contenido de vitamina C y antocianinas durante dos años. El déficit hídrico provocó un aumento en el contenido en antocianinas en todas las accesiones estudiadas, a diferencia de lo ocurrido con la vitamina C, que disminuyó en todos los casos. Esto sugiere que las antocianinas podrían estar implicadas en la tolerancia a sequía en algunas especies de *Lactuca* y, por este motivo, los análisis transcriptómicos se centraron en la búsqueda de genes diferencialmente expresados (DEGs) relacionados específicamente con las antocianinas bajo condiciones de estrés hídrico. Una variedad comercial de lechuga ('Romired') y una especie silvestre (*Lactuca homblei* De Wild) fueron seleccionadas para el estudio transcriptómico (RNA-seq). Entre las accesiones estudiadas, *L. homblei* fue la que mostró el mayor enriquecimiento en el contenido de antocianinas (y la única en la que fue significativo estadísticamente), así como la mayor respuesta a nivel transcriptómico bajo condiciones de estrés. El número de DEGs activados, incluyendo aquellos relacionados con el contenido en antocianinas, fue mayor en *L. homblei*, mientras que el número de DEGs reprimidos fue mayor en 'Romired'. Entre los DEGs que mostraron el mayor cambio en los niveles de expresión y una correlación significativa tanto con el contenido en antocianinas como con el tratamiento de estrés por sequía, se seleccionaron 19 genes candidatos. Cabe destacar que estos genes mostraron expresión diferencial únicamente en *L. homblei*. Además, todos ellos han mostrado funciones relacionadas con el contenido en antocianinas y/o la respuesta a estreses bióticos y abióticos en otras plantas. En las secuencias de la especie silvestre, algunos genes candidatos presentaron polimorfismos con impactos predichos altos o moderados en la función de sus proteínas, como una proteína de choque térmico que participa en el cierre estomático inducido por ácido abscísico, un factor de transcripción involucrado en la respuesta a estreses bióticos y abióticos y una fosfolipasa implicada en la acumulación de antocianinas bajo estrés abiótico. Para validar su expresión mediante PCR cuantitativa (qPCR), se seleccionó previamente el gen de referencia más estable para el experimento de estrés por sequía en base a los datos de RNA-seq. Los patrones de expresión obtenidos con las dos técnicas (RNA-seq y qPCR) fueron consistentes y, por tanto, se pudo confirmar la fiabilidad del estudio transcriptómico. Considerando todos los resultados, *L. homblei* podría ser una fuente valiosa de tolerancia a estrés por sequía y de compuestos beneficiosos para la salud, lográndose también un cultivo de lechuga más sostenible.

GENERAL INTRODUCTION

GENERAL INTRODUCTION

1. Lettuce

1.1. Taxonomy and brief botanical description

Lettuce (*Lactuca sativa* L.) is a leafy vegetable belonging to the Asteraceae family, formerly known as Compositae. It is considered one of the most important families of dicotyledonous plants, being the most evolved and numerous one, with around 24,000 species worldwide (Funk et al., 2009). The genus *Lactuca* is a heterogeneous group that includes the cultivated species, *L. sativa*, and about 100 wild species (Lebeda et al., 2004). Up to 20 of them are considered to belong to the lettuce or primary gene pool (GP1), being *Lactuca serriola* L. the main wild species of the group. The secondary gene pool (GP2) only includes a species, *Lactuca saligna* L., and the tertiary gene pool (GP3) comprises about 11 species, being *Lactuca virosa* L. the most representative member. The groups may change depending on the method used to classify the accessions (Lebeda et al., 2004) and as a consequence of new findings (Wei et al., 2021).

Cultivated lettuce is a predominantly autogamous annual plant, has a basic chromosome number of x = 9 and is a diploid species (2n = 2x = 18). It consists of a very diverse group in which plants are classified in 12 different types based on its morphology: Batavia, Butterhead, Cos, Frillice, Frisée d'Amérique, Gem, Iceberg, Lollo, Multi-divided, Novita, Oakleaf and Stem (UPOV, 2021). There is a great diversity in shape, margin, texture and colour of the leaves among varieties. For example, leaf shape can be circular, lanceolate, oblate and elliptic, among others, and according to the leaf colour, lettuces may be green, semi-red or red. Lettuce leaves grow forming an initial rosette and then either remain loose or cluster close together to form a head. In advanced vegetative stages, the head or cluster of leaves opens to develop a bolting stem up to 1 m long with cauline leaves and small yellow flowers, indicating the start of the reproductive stage. The inflorescences, known as flower heads or capitula, are composed of multiple florets. Each floret produces a single-seeded achene, ribbed and topped with a pappus hair. The seeds can be white, yellow, brown, grey or black, depending on the variety (Mou, 2008).

1.2. Crop origin

Until recently, it has been widely accepted that cultivated lettuce derived from its wild ancestor *L. serriola*, although it is possible that more *Lactuca* spp. may have contributed to some extent. Egypt was initially proposed as the domestication centre (Lindqvist, 1960). In fact, the first depictions of lettuce were found in Egyptian tombs around 2,500 B.C., showing a Cos-like type with leaf scars in the stem. Nevertheless, more hypotheses about the centre of origin have been formulated and other locations have been proposed, such as Mediterranean area, Middle East and Southwest Asia (De Vries, 1997), so the history of lettuce domestication remains to be uncertain [Annex 1 (Medina-Lozano and Díaz, 2022b)]. A recent RNA-seq study reported that the domestication time was around 10,800 years ago (Zhang et al., 2017), whereas a whole-genome resequencing project estimated it around 4,000 B.C. (Wei et al., 2021). In addition, Wei et al. (2021) placed the Caucasus (or close to) as the probable domestication centre of cultivated lettuce. From the Caucasian region, lettuce would have spread to Egypt where oilseed lettuces, the most primitive ones, were used both as vegetables and as oil crop (De Vries, 1997). Then, around 500 B.C., lettuce cultivation was dispersed to Greece and Rome, where Cos lettuce with the entire leaf morphology was originated, likely from crosses with Southern European *L. serriola* accessions (Wei et al., 2021). Later, modern varieties of the crop, like Butterhead lettuces, were emerging in the rest of the European countries over the following thousand years (Lindqvist, 1960; De Vries, 1997). Cultivated lettuce may also have spread from the domestication centre to China between 600 and 900 A.D., where Stem lettuces were obtained through selection around 900 years ago and where they are grown almost exclusively nowadays (De Vries, 1997; Zhang et al., 2017). In America, cultivated lettuce was introduced in the XVI century and breeding procedures led to obtaining the crisphead lettuce (Iceberg) by gene introgression from *L. virosa* (Mikel, 2007).

1.3. Economic interest

Lettuce is one of the major leafy vegetable crops worldwide, mostly consumed raw in salads and sandwiches. It is grown mainly in temperate climates in many countries around the world. Lettuce world production, together with chicory, has increased over a 20% in the last two decades, reaching 27.15 million tonnes in 1.24 million ha in 2022 (FAOSTAT, 2022). The first producer of lettuce and chicory in the world is China, being

responsible for more than half of the total production (55.19%), followed by the USA (12.15%), India (4.28%) and Spain (3.57%) (Figure 1A). Spain is not only the fourth producer of lettuce and chicory in the world, but the leader in Europe, with a total of 969,190 tonnes in 33.540 ha in 2022, which represents more than a quarter of the total European production (Figure 1B). Furthermore, Spain is the leader in terms of exports worldwide, representing around 30% of the total (FAOSTAT, 2022). Lettuce and chicory occupied the second position in the national ranking of vegetable exports in Spain, with more than 790 tonnes exported in 2022, what implied about $960 million in economic value for the country (FAOSTAT, 2022).

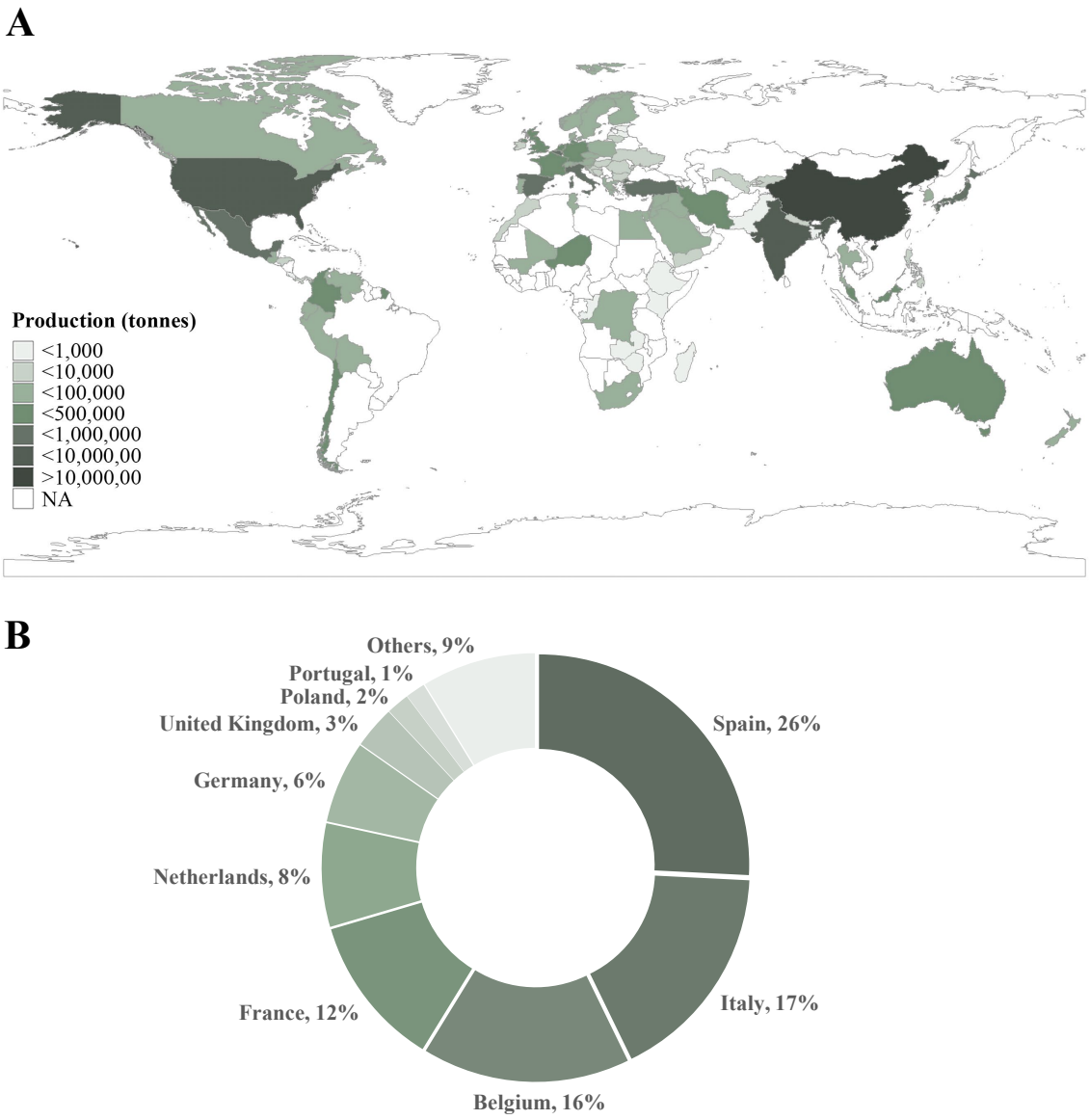


Figure 1. Distribution of lettuce and chicory production in 2022: (A) worldwide (tonnes) and (B) in European countries (%) (FAOSTAT, 2022). NA: not available.

Nevertheless, consumption of lettuce and chicory has decreased by 8.3% in 2022 compared to 2019 in Spain (MAPA, 2022). A likely reason to explain this could be the 3.9% increase in their average price. In Spain, until December 2019, a consistent trend over the years could be observed, when prices increased, consumption of fresh vegetable tended to decrease, what alternated with falls in prices and the expectable rise in consumption (MAPA, 2019). However, since the massive increase in prices during the confinement due to COVID-19 pandemic, price and purchase lines have followed divergent paths, with a sustained rise in the prices while consumption has kept falling.

2. Lettuce breeding

2.1. History

Conventional breeding of lettuce started at the beginnings of the 20th century. In particular, the first breeding programme was initiated in 1922 by the U.S. Department of Agriculture (USDA). In it, several 'Imperial' varieties resistant to brown blight, a fungal disease, were obtained by crossing Butterhead, Cos and Iceberg lettuces (Jagger et al., 1941). After this, many public and private breeding programmes have developed new varieties by genetic recombination through crossing cultivated lettuce varieties. To overcome the limitation of the reduced genetic diversity of *L. sativa*, a predominantly autogamous species, interspecies crosses of the cultivated species with crop wild relatives (CWR) were also implemented in lettuce breeding programmes. The most exploited non-cultivated resources are *L. serriola*, *L. saligna* and *L. virosa*, which have mainly been used in the development of disease resistant lettuces, although in the case of *L. virosa*, that belongs to GP3, previous crosses with *L. serriola* are usually required (Lebeda et al., 2014).

Chemical mutagenesis has also been conducted in lettuce breeding, for example, in the obtaining of two lines with higher seed germination thermotolerance (Huo et al., 2016). Nevertheless, the main and more efficient contributions to the achievement of breeding objectives have been through genomic tools, including the recent publication of the sequence of the lettuce reference genome (Reyes-Chin-Wo et al., 2017). Different types of molecular markers have been used in studies carried out in lettuce, including genetic diversity assessment, linkage map construction and marker-trait association studies. For example, restriction fragment length polymorphisms (RFLPs) were used to

build the first genetic map of lettuce, composed of 53 markers (Landry et al., 1987). Advances in molecular marker discovery allowed to construct an ultra-high-density genetic map of lettuce with almost 14,000 markers using single position polymorphisms (SPPs) (Truco et al., 2013). Other molecular markers used in the above-mentioned studies in lettuce are random amplified polymorphic DNAs (RAPDs) (Tardin et al., 2003), amplified fragment length polymorphisms (AFLPs) (Jansen et al., 2006) and microsatellites or simple sequence repeats (SSRs) (Simko et al., 2009). Nevertheless, all the previous molecular markers have been unarguably ousted by single nucleotide polymorphisms (SNPs) [Annex 2 (Medina-Lozano and Díaz, 2022a)]. In lettuce, SNPs have been obtained using different technologies, from microarrays (Kwon et al., 2012) to next-generation sequencing (Tripodi et al., 2023), and even whole-genome sequencing (Wei et al., 2021). Moreover, there is a list of more than 70K SNPs available at the Lettuce Genome Database (LettuceGDB, https://www.lettucegdb.com). This kind of markers has become essential in marker-assisted selection (MAS) and genomic selection (GS). Both strategies allow to accelerate breeding processes, like in the developing of two Cos lettuce varieties resistant to the dieback disease with good shelf life (Simko, 2013).

Finally, bred lettuces have been developed through different genetic engineering approaches like transgenesis and CRISPR/Cas9-mediated genome editing, being the latter the most innovative technique to date. The first report on the production of transgenic lettuces was about the development of a transformation system mediated by *Agrobacterium tumefaciens* as part of a long-term strategy aimed at increasing disease resistance (Michelmore et al., 1987). Woo et al. (2015) obtained genome-edited lettuces for the first time by knocking out a gene that encodes a negative regulator in the brassinosteroid signalling pathway, generating plants theoretically indistinguishable from those generated by traditional breeding practices.

Nowadays, more than 11,000 entries of lettuce-related accessions can be found in the genebanks distributed worldwide, including improved cultivars, research material, traditional varieties and landraces, and wild species (Genesys, 2024). Nevertheless, the largest contribution to lettuce breeding has been made by the private sector. Public crop breeding programmes have decreased in the last years and 95% of the vegetable market was controlled by just five seed companies in 2012, including Bayer, Limagrain, Monsanto (acquired by Bayer in 2018), Rijk Zwaan and Syngenta (Mammana, 2014). In

particular, about 80% of the European-wide plant breeders' rights for lettuce are held by four of the largest corporations, including three of those just mentioned (Limagrain, Rijk Zwaan and Syngenta) along with Enza Zaden (OECD, 2018). Therefore, the real number of lettuce accessions is likely to be higher, as detailed information about private breeding programmes is not usually published for obvious economic reasons (Collard and Mackill, 2008).

2.2. Targets in the breeding programmes: past, present and future

Since the beginning of lettuce breeding, main efforts have been focused on developing disease resistance varieties and on increasing the crop yield. When looking at the literature, a combined search for "lettuce", "breeding" and "disease" terms finds 1,284 hits in the Web of Science database (https://www.webofscience.com/), and 634 hits for "lettuce", "breeding" and "yield" (July 2024). In less extent, lettuce breeding interest has also been directed to agronomic (bolting time, leaf shape and colour) and postharvest (shelf life) characters. Flavour is another trait that has been improved since its domestication started with the aim of reducing the bitterness provided by CWR (De Vries, 1997). Selection against it has caused a detrimental effect on lettuce health-promoting properties, for example through the reduction of sesquiterpene lactones, which have potential benefits to human health, but are responsible for the bitter taste (van Treuren et al., 2018). In addition, nutritional value together with tolerance to abiotic stresses have been less exploited in lettuce breeding programmes: combined searches of "lettuce" and breeding" plus "nutritional value" or "abiotic stress" yielded 110 and 111 hits, respectively, as pointed out before (Damerum et al., 2020; Simko et al., 2021).

When looking for information about lettuce breeding in the private sector, a similar pattern to the observed in the public sector or in academia is found. In the catalogues or news published by the largest seed companies, the announced new varieties are generally resistant to one or more diseases, mainly those caused by *Bremia* and *Fusarium*, the major fungal agents affecting lettuce. Characters like yield, taste, morphology, colour and shelf life are also frequent objectives for the companies. New varieties tolerant to abiotic stresses are less common, being heat the main target stressor, since lettuce is a cool season crop sensitive to high temperatures. In this sense, the main goals are to obtain varieties with delayed bolting and with low susceptibility to tip burn (not exactly caused by high temperatures but aggravated by them). However, climate

change is causing a reorientation of breeding objectives, for instance, to ensure the availability of lettuces during the whole harvest calendar. Finally, nutritional value is not among the priorities for the private sector, so it is difficult to find new varieties with enhanced nutritional content in the catalogues of the main companies.

In spite of being up to 97% water (Mou, 2005), lettuce contains important nutrients and other phytochemicals, such as vitamins, fibre, carotenes and polyphenols, known to provide benefits for human health (Llorach et al., 2008; Kim et al., 2016). Indeed, thanks to its medicinal properties, provided by some of its bioactive compounds, lettuce has traditionally been used in the treatment of disorders like insomnia, dry cough, rheumatic pain and even anxiety (Hassan et al., 2021). However, lettuce is not especially nutritious since some compounds are present in low quantities, particularly when compared to other leafy vegetables such as chard, spinach, cabbage or watercress (USDA, 2022). In addition, its nutritional composition depends highly on the genotype. Differences in both primary and secondary metabolite contents are found among the most common commercial varieties (USDA, 2022). It has also been reported that wild relative species are richer than commercial varieties in some phytochemicals like vitamin C (van Treuren et al., 2018), some phenolic compounds (Damerum et al., 2015) and carotenoids (Mou, 2005). On one hand, this reflects the high improvement margin in terms of the content in nutrients and health beneficial compounds existing in the crop. This, along with lettuce being one of the most popular vegetables among consumers all around the world, makes it the ideal candidate for undertaking a breeding programme aimed at enriching its content of nutrients and bioactive compounds. Moreover, making healthier and more accessible food is a task of the utmost importance in the current context, where malnutrition affected 9.2% of the world population in 2022, mainly in developing countries, but also in developed ones. In this last case, it is due to micronutrient deficiencies (known as “hidden hunger”) and not to food availability, being the causes very different, such as inadequate weight loss diets, excessive alcohol consumption and drug abuse (FAO et al., 2023).

On the other hand, the scenario of climate change in which the planet is immersed makes it necessary a deeper exploitation of breeding for tolerance to abiotic stresses, due to the unprecedent levels of negative effects on agriculture caused by adverse environmental conditions (FAO, 2023). In addition, it is known that some abiotic stresses

cause an increase in phytochemical content of some crops, including lettuce (Qaderi et al., 2023). Therefore, the development of new lettuce varieties more resilient to climate change and at the same time healthier is a feasible possibility that would result beneficial for both farmers and consumers.

2.3. Genetic resources: lettuce wild relatives and traditional varieties

The genetic diversity of cultivated lettuce, already limited due to its predominantly autogamous nature, has resulted even more narrowed down during the adaptation and selection processes throughout its domestication. CWR are valuable resources that harbour great genetic variability, essential for lettuce improvement. In fact, they have often been included in breeding programmes, being *L. serriola*, *L. saligna* and *L. virosa* the wild relatives most widely used, as mentioned before, and also the most represented in genebank collections (Lebeda et al., 2014). Nevertheless, there are many available options apart from them, since *Lactuca* gene pool consists of about 100 wild species. In comparison to cultivated lettuce, its wild relatives have shown more resistance to different diseases (Lebeda et al., 2014) and better nutritional values for specific compounds, such as vitamin C (van Treuren et al., 2018) and some polyphenols (Damerum et al., 2015), what makes evident their valuable contribution to *L. sativa* improvement. However, they have two main disadvantages. Firstly, not all the *Lactuca* spp. are sexually compatible with the cultivated lettuce. Having said that, members of GP1 are generally interfertile, which facilitates classical breeding; otherwise, interspecific crosses could be assisted by some techniques, such as embryo rescue, or strategies like bridge crosses. The second inconvenient is the called linkage drag, that is, genes that negatively affect crop performance are inherited along with the genes of interest because of their physical proximity and/or strong linkage disequilibrium (LD), being difficult to separate them by conventional techniques like breeding crosses. As a consequence, the use of lettuce wild relatives may have been limited in breeding programmes to avoid their undesirable traits, such as bitter taste, leaf prickles or early bolting.

Additionally, traditional varieties or landraces are also valuable resources to be used in breeding programmes. In terms of domestication and/or breeding, they are something intermediate between the two extremes of the wide spectrum of plant material: CWR and commercial varieties, probably offering advantages over both of them. On one hand, both sexual incompatibility barriers and undesirable traits of CWR transferred by

linkage drag are avoided. On the other hand, landraces and traditional varieties harbour higher genetic variability than commercial varieties. Moreover, there are a plethora of studies in which traditional varieties and landraces have been reported to be richer in nutrients and phytochemicals when compared to commercial varieties in different crops [Annex 3 (Medina-Lozano and Díaz, 2021)].

According to the biological status of accessions, traditional varieties and landraces account for more than 30% of total plant accessions registered in genebanks worldwide, excluding the "not specified" material (Genesys, 2024). Thus, they are the most represented group, ranking above breeding/research material, advanced/improved cultivars and CWR, what highlights their importance. In the particular case of lettuce, traditional varieties and landraces account for 21% of the more than 11,000 crop-related accessions registered in Genesys (2024), only behind advanced/improved cultivars. However, little attention has been paid to them in lettuce breeding programmes. Recently, lettuce traditional varieties and landraces are starting to be used in the assessment of nutrient and bioactive compound content in several studies (Martínez-Ispizua et al., 2022; Mallor et al., 2023; Zeljković et al., 2023). Even if not very often, they have also been used in association studies of disease resistance, agronomic traits and anthocyanin content (Moreno-Vázquez et al., 2003; Wei et al., 2021; Tripodi et al., 2023).

3. Lettuce nutrients and bioactive compounds: vitamin C and anthocyanins

Lettuce is a widely consumed crop, ranking among the most popular vegetables worldwide. It is perceived as a healthy food, so its high consumption may be explained by the global growing awareness of the need of having healthier diets in recent years. In addition, wellness-conscious consumers are prioritising more and more natural products over processed ones and, indeed, lettuce is mainly consumed raw, except in some cases where stems or leaves are cooked, especially in China. Nevertheless, the most commercialized and therefore the most consumed lettuce varieties have a low nutritional value. Even so, due to its high consumption, lettuce is an important source of different phytochemicals including vitamins and phenolic compounds. Among the vitamins, vitamin C is the predominant one, although vitamins A, some of the B-group, K and E can be also found in lettuce (USDA, 2022). It also contains some phenolic compounds, such as chlorogenic and caffeic acids, quercetin and kaempferol derivatives, anthocyanins, etc. Anthocyanins are important compounds in red-leaf lettuces and both

vitamin C and anthocyanins have powerful antioxidant properties that provide health benefits for humans and protection against stresses for plants.

3.1. Vitamin C

Vitamin C is a water-soluble vitamin that represents one of the most important indicators of food nutritional quality. It is an essential nutrient for humans which has to be taken through the diet to ensure a normal metabolic functioning because humans have lost the capacity to synthesize it (Jaffe, 1984). The importance of vitamin C for humans was brought to light in the mid-18th century, when it was discovered that fresh citrus juice cured scurvy (Lind, 1753). Besides, in the first half of the 20th century, it was demonstrated that a water-soluble vitamin identified as vitamin C helped to prevent that same disease (Svirbely and Szent-Györgyi, 1932). Nowadays, vitamin C is known to participate in several processes in the human body, including collagen biosynthesis, reduction of cholesterol, inorganic iron absorption as well as having potential preventive roles in cardiovascular diseases and cancer thanks to its antioxidant capacity (Li and Schellhorn, 2007). More recently, a potential role of vitamin C in managing symptoms of respiratory viral infections, including COVID-19, has been described (Fath et al., 2022).

Vitamin C is composed of ascorbic acid (AA) and dehydroascorbic acid (DHAA) (Figure 2A), being AA the molecule with the highest biological activity. AA biosynthesis has been well-characterized in plants. The main route is the D-mannose/L-galactose pathway, also known as Smirnoff-Wheeler pathway (Wheeler et al., 1998), though three more alternative pathways have been identified to date: L-glucose, myo-inositol and D-galacturonic acid routes (Venkatesh and Park, 2014). Apart from biosynthesis processes, AA levels are also maintained through its transportation, degradation and recycling. AA and DHAA are easily interconvertible either spontaneously or enzymatically, for instance, through the ascorbate-glutathione cycle in which AA is oxidised to DHAA by the ascorbate peroxidase (APX) and recycled from DHAA by the dehydroascorbate reductase (DHAR) (Figure 2) (Apel and Hirt, 2004). This indicates that DHAA may act as a reservoir of vitamin C, contributing to maintain the AA pool. Besides, DHAA has some biological activity itself, what highlights its importance although it has often been neglected in vitamin C studies probably because its direct quantification is difficult due to its low absorptivity in the ultraviolet (UV) range of the spectrum.

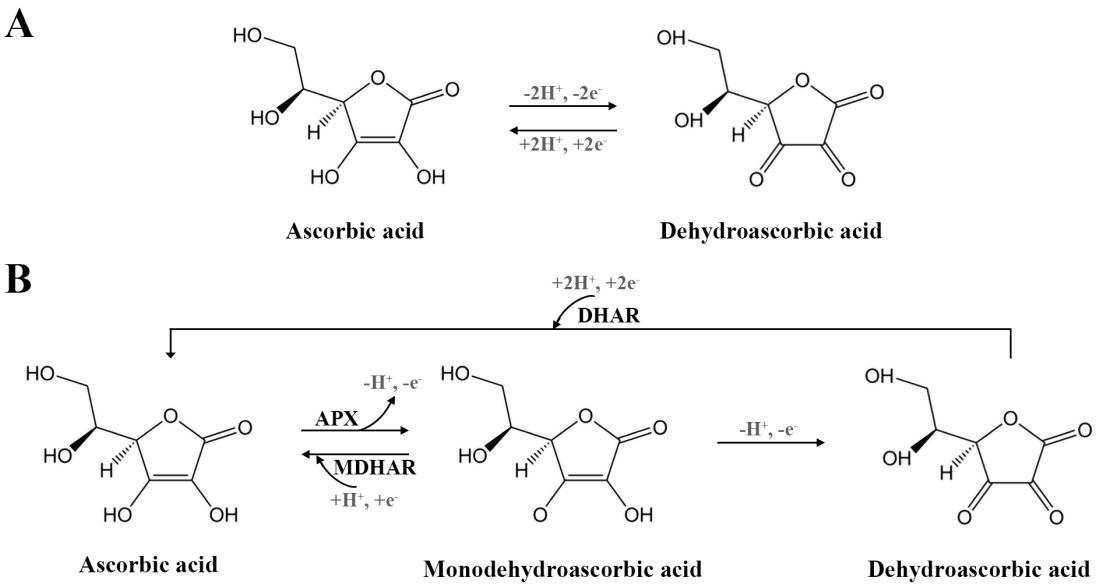


Figure 2. Interconversion of ascorbic acid and dehydroascorbic acid (A) spontaneously and (B) enzymatically through monodehydroascorbic acid in the oxidation reaction. APX: ascorbate peroxidase, MDHAR: monodehydroascorbate reductase, DHAR; dehydroascorbate reductase. Edited from Apel and Hirt (2004).

In plants, vitamin C is a cofactor for several enzymes, participates as cell signalling modulator in different processes, such as cell wall growth and cell division and expansion, and is involved in photosynthesis and respiration (Ishikawa et al., 2006). In lettuce, until the beginning of this thesis, vitamin C content had been assessed in only a few commercial varieties, as well as in CWR (Llorach et al., 2008; van Treuren et al., 2018). Regarding the improvement of vitamin C content in lettuce, it has been achieved through different methods, such as conventional temporary fortification with iodine supplementation (Dyląg et al., 2023), changes in growing conditions like different light intensities (Zhou et al., 2007), or modern genetic engineering approaches, including both transgenesis (Kim et al., 2004; Guo et al., 2013) and genome editing (Zhang et al., 2018a). Nevertheless, those conventional approaches present the disadvantage of providing only temporary effects and genetic engineering techniques must face legislative obstacles in some countries. To the best of our knowledge, neither conventional breeding aimed at enhancing vitamin C content nor marker association studies with this trait have been carried out in lettuce up to today.

3.2. Anthocyanins

Anthocyanins are a group of water-soluble phenolic compounds that are widely found in the plant kingdom, and have also been recently identified in some fungi (Bu et

al., 2020). Many studies have demonstrated anthocyanin preventive and therapeutic roles on human health thanks to their powerful antioxidant properties. They are known to protect against diseases like diabetes, obesity, cancer and cardiovascular and neurological pathologies (Yousuf et al., 2016). Anthocyanins are one of the most important water-soluble pigments in plants, responsible for orange, red, purple and blue colours of leaves, fruits, petals and seeds, together with other compounds. These colours make lettuce more appealing for consumers and also reveal the presence of health-promoting compounds, positively influencing their buying behaviour. This positive influence has been proved in other crops, like cactus pear fruit (Migliore et al., 2015) and pepper (Di Vita et al., 2024).

Anthocyanins are flavonoids that consist of an aglycon backbone (anthocyanidin) and one or more glucoside conjugates (monosaccharides like glucose, rhamnose, galactose, arabinose and xylose, in descending order of abundance, disaccharides like rutinose, or even trisaccharides), and these, in turns, can often be acylated with acids (e.g., acetic or malonic acid) in different positions (Figure 3A). More than 650 anthocyanins have been identified to date, most of them derived from the six most common anthocyanidins: cyanidin, delphinidin, malvidin, pelargonidin, peonidin and petunidin (Figure 3B) (Andersen and Jordheim, 2010). Anthocyanins are synthesized through the general phenylpropanoid pathway followed by the specific flavonoid pathway. Enzymes involved in the biosynthesis pathway as well as transcription factors that regulate them, both positive and negatively, have been extensively studied, as reviewed in Chaves-Silva et al. (2018). Specifically in lettuce, structural anthocyanin-related genes and some transcription factors have also been identified (Zhang et al., 2016; Su et al., 2020).

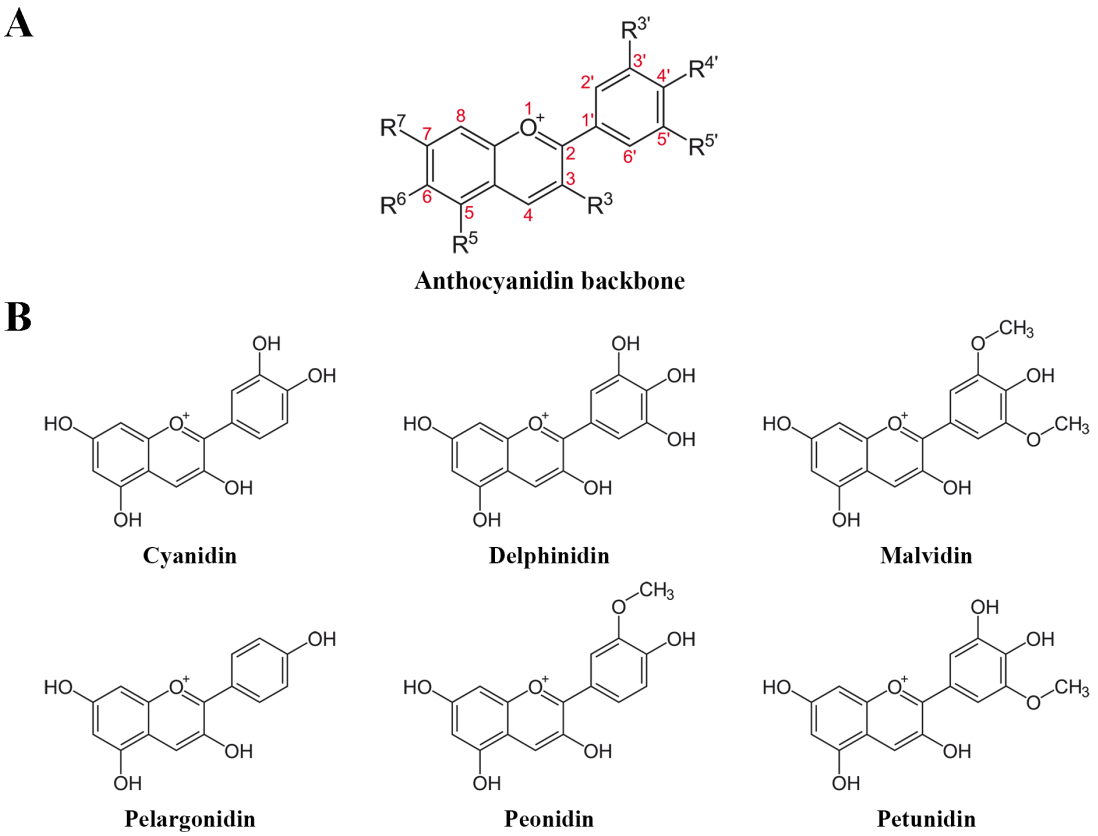


Figure 3. Chemical structure of (A) the anthocyanidin backbone and (B) the six most common anthocyanidins: cyanidin, delphinidin, malvidin, pelargonidin, peonidin and petunidin. Red numbers: positions of the benzopyrylum (n) and phenyl (n') rings. R^i: substituents positions of anthocyanidins, most commonly R^3: H, OH, OGly; R^5, $R^{3'}$ and $R^{5'}$: H, OH, OCH_3, OGly; R^6: H, OH; R^7 and $R^{4'}$: OH, OCH_3; Gly: glycosyl residue. Edited from Chaves-Silva et al. (2018).

In plants, anthocyanins play a role not only in reproduction by attracting animals for seed and pollen dispersal thanks to their eye-catching colours (Tanaka et al., 2008), but also in protection during vulnerable developmental stages (i.e., early and senescent stages) and against radiation and several biotic and abiotic stresses due to their antioxidant properties (Landi et al., 2015). In lettuce, the identified anthocyanins are mainly cyanidin glucosides (Wu and Prior, 2005; Viacava et al., 2017) and less frequently peonidin and delphinidin glucosides (Baek et al., 2013), which are responsible for the leaf colour in red and semi-red varieties.

Anthocyanin content has been analysed mainly in commercial varieties and more rarely in traditional varieties and CWR (Kim et al., 2016; Wei et al., 2021). Accumulation of anthocyanins has been reported when lettuce is subjected to different growing conditions, including different light wavelengths (Sng et al., 2021), UV irradiation (Tsormpatsidis et al., 2008) and cold (Becker et al., 2014). As far as we know, genetic

engineering techniques have not been used to enhance anthocyanin content in lettuce, but several genome-wide association studies (GWAS) have targeted these compounds (Kwon et al., 2013; Wei et al., 2021; Tripodi et al., 2023), even though they were assessed as agronomic characters responsible for leaf colour and not because of their beneficial properties both for human health and for the plant performance itself.

4. Abiotic stress: drought

At present, agriculture is being seriously threatened by the adverse environmental conditions derived from climate change, whose negative consequences are of higher magnitude and happening earlier than projected (IPCC, 2023). In particular, droughts are one of the most worrying abiotic stresses and their durations are expected to increase drastically if global temperature rises as predicted. Drought stress affects different physiological and biochemical processes of the plant (Farooq et al., 2009). Nevertheless, controlled water deficit may have a positive impact by increasing some antioxidant compounds with health-promoting properties, as it has been reported in different crops (Qaderi et al., 2023). This might be due to the activation of metabolite-mediated plant stress responses, for instance, through antioxidant molecules which may neutralise oxidative stress, in particular, reactive oxygen species (ROS), commonly induced by abiotic stresses, among others. Anthocyanin accumulation under drought stress seems to be a widely spread response, as it has been demonstrated in crops like grapevine (Ju et al., 2019), hibiscus (Hinojosa-Gómez et al., 2020) and strawberry (Rugienius et al., 2021). By contrast, plant response to water deficit in terms of vitamin C content has not been so extensively studied, despite its importance for both plants and human health, and the results are not all in the same direction as in the case of anthocyanins. So, both increases (Sarker and Oba, 2018; Niyazova and Huseynova, 2024) and decreases (Seminario et al., 2017) of vitamin C content as a consequence of drought stress have been reported in different crops. In lettuce, accumulation of both anthocyanin and vitamin C has been observed in response to abiotic stresses other than drought, including changes in light intensities, UV irradiation and cold, as mentioned in the previous section. Regarding drought stress response in the crop, vitamin C content has been reported to increase (Koyama et al., 2012) and also to decrease (Zeljković et al., 2023) in studies carried out with different cultivated lettuces, whereas anthocyanins have barely been studied. Therefore, more research is needed to understand the lettuce response to water deficit

mediated by these compounds, especially when seeing the variability in the behaviour of vitamin C within lettuce and in different crops as well as the lack of studies in the case of anthocyanins. This research is particularly important given that drought consequences might be devastating for agriculture in the near future.

5. Omic approaches in lettuce studies

Development of new technologies has been key for plant breeding, especially in recent years with the advances in omic approaches, mainly in genomics, transcriptomics, proteomics, epigenomics and metabolomics. Among them, genomic tools are likely the most used in lettuce to date, which have been favoured by the discovery of SNPs and the development of molecular markers and especially by the publication of a reference genome in the species (Reyes-Chin-Wo et al., 2017). There is a multitude of genomic tools at hand to improve crops, for instance, enhancing their nutritional quality, as reviewed in Medina-Lozano and Díaz (2022a) (Annex 2). Nevertheless, their use in lettuce has mainly been focused on disease resistance and agronomic traits, as highlighted before in this introduction.

Transcriptomic studies targeting vitamin C and anthocyanins in lettuce have often been directed to study the effects of UV irradiation (Zhou et al., 2023) and different light conditions (Zhang et al., 2018b; Wada et al., 2022), respectively. In the case of anthocyanins, differences between green and red varieties have also been assessed (Moreno-Escamilla et al., 2020; Su et al., 2020). To the best of our knowledge, only an RNA-seq study has been conducted until now to study drought stress in lettuce (Koyama et al., 2021), in which the observed increase in antioxidant compounds was likely attributed to a higher antioxidant enzyme activity, though no induction of antioxidant enzyme-related genes was detected.

Recently, metabolomic approaches have been used in lettuce with different aims. An unprecedent large scale metabolomic study through an LCMS (liquid chromatography/mass spectrometry) untargeted analysis was conducted by van Treuren et al. (2018). In this study, anthocyanins and vitamin C were assessed in different gene pools, including accessions of commercial and primitive lettuces and wild relatives. The authors reported that phytochemical profiles appeared to be species-specific. Furthermore, vitamin C and anthocyanins have been widely quantified using less complex

techniques such as liquid chromatography with diode array detectors and spectrophotometric methods, for example in Santos et al. (2016) and Chen et al. (2019), where changes in the amounts of both compounds were reported when lettuce plants were cultivated using different food-based composts or were subjected to UV radiation, respectively.

Finally, studies of proteomics, phenomics and epigenomics have also been conducted in lettuce. For instance, Hao et al. (2018) found proteins associated with photosynthesis and auxin metabolism related to high temperature-induced bolting. Adhikari et al. (2019) used phenomics to conduct reproducible, fast and non-destructive analyses of phytochemicals under salt stress in a large number of plants. Lastly, Cao et al. (2024) found increased levels of DNA methylation as a consequence of lettuce domestication by carrying out epigenomic studies.

Large scale data coming from studies applying all these technologies can be already found in the LettuceGDB. In addition, germplasm resources and tools for data exploration are accessible, which are part of the ongoing efforts that hopefully will improve the database (Guo et al., 2023). The possibility to integrate these technologies in a multi-omic approach, combined with bioinformatic tools, opens up a plethora of opportunities to obtain enhanced crops in terms of yield, disease resistances, tolerance to abiotic stresses, nutritional quality, etc., besides leading us to more efficient and sustainable agricultural practices in relatively short periods of time. To obtain lettuce varieties that are both more resilient and richer in compounds with health benefits is important to explore the phytochemicals of interest under control and stress conditions in different accessions. Additionally, it is necessary to combine different type of studies, such as genomics and transcriptomics. The next step would involve conducting physiological analyses, which are indispensable to ensure that plant yield is maintained while improving the desired traits.

6. Objectives

The general objective of this thesis is multiple. First, it aims at the selection of a lettuce variety with enhanced content of antioxidants (vitamin C) and, second, it seeks to investigate the effect of drought stress on antioxidant compounds (vitamin C and anthocyanins) in lettuce-related germplasm, delving into its molecular bases when the stress causes an increase in their content. To achieve them, the following specific objectives have been designed:

Objective 1. Quantification of vitamin C and anthocyanins in different lettuce plant material including commercial and traditional varieties as well as wild relatives (*Lactuca* spp.).

Objective 2. Selection of the variety with the highest vitamin C content and genetic characterization along with GWAS of vitamin C content in cultivated germplasm (*L. sativa*).

Objective 3. Evaluation of drought stress impact on vitamin C and anthocyanin contents in lettuce varieties and wild relatives.

- Objective 3.1. Vitamin C and anthocyanin quantification in plants subject to control and drought stress conditions.
- Objective 3.2. Transcriptomic study of the increase in anthocyanin content in response to water deficit and search for putative causative polymorphisms in candidate genes.

CHAPTER 1

Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content

Medina-Lozano, I, Bertolín, JR, Díaz, A (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* 359, 129864. https://doi.org/10.1016/j.foodchem.2021.129864.

CHAPTER 1

Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content

Justification

Vitamin C and anthocyanins are two of the most important antioxidant compounds in lettuce. First, vitamin C is one of the main indicators of nutritional quality in plant foodstuffs and is the predominant vitamin in lettuce. Second, anthocyanins are bioactive compounds that provide potential benefits for human health as well as being responsible for the leaf colour of red varieties. There are very few studies focused on improving the content of nutrients and health-promoting phytochemicals in lettuce. Considering, on one hand, the growing interest of consumers in healthier diets and, on the other hand, the relevance of vitamin C and anthocyanins for both crop performance and human wellbeing, it becomes essential to analyse them as the starting point of any breeding programme aimed at developing lettuce varieties enriched with any of those compounds. Not only wild relative species but also traditional varieties are valuable resources for crop breeding, so both were included in this study together with commercial varieties. As far as we know, this is the first time that both vitamin C and anthocyanins have been analysed in such a diverse *Lactuca* germplasm.

Before starting the vitamin C assessment, it was necessary to develop an optimized protocol that ensured an accurate and efficient quantification from a high number of samples in a short period of time due to its high instability (it can start degrading just 4 h after extraction). That protocol to quantify, not only AA, the main form of vitamin C, but also DHAA, which also shows biological activity, has been published as a method paper and is included in Annex 4 (Medina-Lozano et al., 2020).

Food Chemistry 359 (2021) 129864

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin C and anthocyanin content

Inés Medina-Lozano [a,c], Juan Ramón Bertolín [b,c], Aurora Díaz [a,c,*]

[a] Unidad de Hortofruticultura, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Avda. Montañana 930, 50059 Zaragoza, Spain
[b] Unidad de Producción y Sanidad Animal, Centro de Investigación y Tecnología Agroalimentaria de Aragón, Avda. Montañana 930, 50059 Zaragoza, Spain
[c] Instituto Agroalimentario de Aragón – IA2 (CITA-Universidad de Zaragoza), Zaragoza, Spain

ARTICLE INFO

Keywords:
Lettuce
Crop wild relatives
Vitamin C
Ascorbic acid
Dehydroascorbic acid
Anthocyanins
Germplasm
UPLC-UV

ABSTRACT

Lettuce is the most consumed leafy vegetable though the most popular varieties have a low nutritional value. Our objective was to accurately quantify vitamin C and anthocyanins in wild relatives, and commercial and traditional varieties.

Wild species and traditional varieties contained more total ascorbic acid (TAA) than commercial varieties (21% and 8%, respectively). In contrast, commercial varieties had significantly higher content of anthocyanins than traditional varieties and wild species (6 and 8 times more, respectively). TAA was significantly higher in green than in red lettuces (18%). TAA was also significantly higher in the leaves of two wild species than in stems. Cyanidin 3-*O*-(6′-*O*-malonylglucoside) was the most abundant anthocyanin (97%), present in most samples.

The rankings of accessions by vitamin C and anthocyanin contents can be useful for consumers worried about the impacts of food on their wellbeing and for breeders aiming to improve lettuce by biofortification with health-promoting compounds.

1. Introduction

Lettuce (*Lactuca sativa* L.) is the most consumed and the most cultivated leafy vegetable worldwide with production continuing to increase yearly. It was estimated that the total area harvested, globally, was more than 1.27 million hectares in 2018, with a total production of approximately 27.3 million tonnes (FAOSTAT, 2018).

L. sativa belongs to Asteraceae family and varieties are classified according to their morphology in 12 types (UPOV, 2019): Batavia, Butterhead, Cos, Frillice, Frisée d'Amérique, Gem, Iceberg, Lollo, Multidivided, Novita, Oakleaf and Stem; and colour: green, semi-red and red. Wild relatives within the genus *Lactuca* also demonstrate enormous variability with more than 100 species (Lebeda et al., 2004). In addition to preserving this rich patrimony, many germplasm banks are carrying out a detailed evaluation of the plant material. The main efforts have been focused on morphological and, more recently, genetic characterisation (Mallor & Díaz, 2016), while the metabolic aspects have barely been taken into account. In the case of lettuce, this metabolic characterisation is very important because the crop is a source of phytonutrients with health-promoting properties, such as phenolic compounds and vitamins (Llorach et al., 2008). However, the most consumed varieties worldwide happen to have a low nutritional value. Differences among the groups under study (commercial varieties, traditional varieties and wild relatives) in the concentration of two compounds with a proven beneficial effect on human health, vitamin C and anthocyanins, will be sought out. In recent years, there has been a growing interest among consumers in the impact of food on health. In fact, many of them are willing to pay a higher price for healthier products, with a higher nutritional value or with potential benefits in disease prevention. This could represent an added value for traditional varieties, as well as a very interesting breeding objective.

Vitamin content, especially vitamin C, is one of the most relevant indicators of the nutritional quality of fruits and vegetables. Vitamin C can be found in two interconvertible forms, ascorbic acid (AA), that exhibits antioxidant activity, and dehydroascorbic acid (DHAA), its oxidation product (Lee & Kader, 2000). Vitamin C is an essential micronutrient for humans, required to ensure a normal physiological function, which has to be taken as part of the diet because humans have

* Corresponding author at: Unidad de Hortofruticultura. Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Avda. Montañana 930, 50059 Zaragoza, Spain.
E-mail address: adiazb@cita-aragon.es (A. Díaz).

https://doi.org/10.1016/j.foodchem.2021.129864
Received 22 August 2020; Received in revised form 5 April 2021; Accepted 11 April 2021
Available online 20 April 2021

lost the capacity to synthesize it (Carr & Frei, 1999). Its beneficial properties are related to its role in different processes in the human body, such as collagen formation, reduction of cholesterol levels, inorganic iron absorption, and most important of all, enhancement of the immune system due to its antioxidant activity (Carr & Frei, 1999).

Other compounds with strong antioxidant activity are anthocyanins, a class of water-soluble phenolic compounds which until recently were thought to be synthesized exclusively by plants but have also been found in some fungi (Bu et al., 2020). These play an important role in human health by preventing neuronal and cardiovascular diseases, reducing diabetes risk, and exhibiting anti-cancer activity (Yousuf et al., 2016). In lettuce, they are responsible for the red colour of the leaves, which is an important characteristic, not only because it has a major influencing effect on consumer buying behaviour, but also because it reveals the

Table 1
Description of the plant material used in the present study, commercial and traditional lettuce varieties, as well as wild relatives (*Lactuca* spp.).

Accession name	Species	Group	Type[a]	Leaf colour	Lettuce gene pool	Origin	Source[b]	Accession number
'Begoña'	*Lactuca sativa* L.	Commercial variety	Batavia	Green	Primary	Spain	Ramiro Arnedo Semillas S.A.	–
'Dolomiti G12'	*Lactuca sativa* L.	Commercial variety	Gem	Green	Primary	Spain	Ramiro Arnedo Semillas S.A.	–
'Likarix'	*Lactuca sativa* L.	Commercial variety	Frisée d'Amérique	Red	Primary	Netherlands	CGN	CGN24522
'Lollo Rosso'	*Lactuca sativa* L.	Commercial variety	Lollo	Red	Primary	Italy	CGN	CGN09385
'Winter Crop'	*Lactuca sativa* L.	Commercial variety	Butterhead	Green	Primary	Hungary	BGHZ	CGN05853
'Nestorix'	*Lactuca sativa* L.	Commercial variety	Lollo	Red	Primary	Netherlands	CGN	CGN24712
'Red Sails'	*Lactuca sativa* L.	Commercial variety	Lollo	Red	Primary	Germany	CGN	CGN19014
'Revolution'	*Lactuca sativa* L.	Commercial variety	Lollo	Red	Primary	Netherlands	CGN	CGN20714
'Romana Inverna'	*Lactuca sativa* L.	Commercial variety	Cos	Green	Primary	Spain	BGHZ	BGHZ3604
'Romired'	*Lactuca sativa* L.	Commercial variety	Lollo	Red	Primary	Netherlands	CGN	CGN24713
'Lechuga de Beceite'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ2006
'Lechuga de Bureta'	*Lactuca sativa* L.	Traditional variety	Cos	Semi-red	Primary	Spain	BGHZ	BGHZ4927
'Lechuga de Ensalada'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ2031
'Lechuga de Híjar'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ0529
'Lechuga de Subías'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ1852
'Lechuga del Pirineo'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ2229
'Lechuga del Valle de Tena'	*Lactuca sativa* L.	Traditional variety	Butterhead	Green	Primary	Spain	BGHZ	BGHZ1850
'Lechuga Romana Zaragozana'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ4306
'Lengua de Buey'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ2004
'Morada de Belchite'	*Lactuca sativa* L.	Traditional variety	Cos	Semi-red	Primary	Spain	BGHZ	BGHZ0527
'Morada de Bernués'	*Lactuca sativa* L.	Traditional variety	Batavia	Semi-red	Primary	Spain	BGHZ	BGHZ2097
'Morada de Sorripas'	*Lactuca sativa* L.	Traditional variety	Cos	Semi-red	Primary	Spain	BGHZ	BGHZ2026
'Oreja de Mulo'	*Lactuca sativa* L.	Traditional variety	Cos	Green	Primary	Spain	BGHZ	BGHZ0532
Lactuca biennis	*Lactuca biennis* (Moench) Fernald	Wild crop relative	–	Dark green	Tertiary	North America	BGHZ	BGHZ4761
Lactuca dentata	*Lactuca dentata* Makino	Wild crop relative	–	Green	Tertiary	East Asia	BGHZ	BGHZ4345
Lactuca dregeana	*Lactuca dregeana* DC.	Wild crop relative	–	Dark green (red stems)	Secondary	South Africa	BGHZ	BGHZ3670
Lactuca floridana	*Lactuca floridana* L. Gaertn.	Wild crop relative	–	Dark green	Tertiary	North America	BGHZ	BGHZ5323
Lactuca homblei	*Lactuca homblei* De Wild	Wild crop relative	–	Green (red nerves)	Tertiary	Central and South Africa	BGHZ	BGHZ5322
Lactuca squarrosa	*Lactuca squarrosa* (Thunb.) Miq.	Wild crop relative	–	Dark green (red stems)	Tertiary	Southeast Asia	BGHZ	BGHZ5124
Lactuca virosa	*Lactuca virosa* L.	Wild crop relative	–	Green (red nerves)	Secondary[c] or tertiary[d]	East and North Africa	BGHZ	BGHZ4051

[a]According to UPOV (2019).
[b]BGHZ: Vegetable Germplasm Bank of Zaragoza (Spain); CGN: Centre for Genetic Resources (Wageningen, Netherlands).
[c]According to Koopman et al. (1998) and Koopman (1999).
[d]According to Zohary (1991).

presence of beneficial compounds to health.

Both, vitamin C and anthocyanins, not only have proven beneficial effects on human health but they also help the plants to cope with stresses (biotic and abiotic) mainly thanks to their antioxidant properties, and, in the case of some anthocyanins, protecting the leaves from high radiation without compromising photosynthesis, what ultimately renders higher yields.

On one hand, it is evident that wild *Lactuca* spp. resources have a potential interest in breeding programmes (Lebeda et al., 2014). However, the efforts have been mainly directed to incorporate resistance to pests and diseases, and few studies are focused on enhancing the nutritional value of the crop (van Treuren et al., 2018). Furthermore, only a few wild species have been used in lettuce breeding programmes, even if the diversity within the wild *Lactuca* spp. is worthy of consideration. On the other hand, studies such as those carried out by Llorach et al. (2008) and by Mulabagal et al. (2010), in which anthocyanins and vitamin C were quantified in some green and red varieties, are of great interest. However, they are only focused on commercial varieties. In contrast, the main objective of this study is the analysis and quantification of the nutritional value, specifically, the vitamin C and anthocyanin contents, not only of commercial varieties, but also traditional varieties and lettuce wild relatives.

2. Materials and methods

2.1. Reagent and standards

Acetic acid (≥99% purity), hydrochloric acid (37% purity), *meta*-phosphoric acid (MPA, 33.5–36.5% purity) and sulfuric acid (95–98% purity) were purchased from Sigma-Aldrich (Madrid, Spain). 1,4-Dithiothreitol (DTT, ≥98% purity) and 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris, ≥99.9% purity) were provided by Roche (Madrid, Spain). Ethylenediaminetetraacetic acid disodium salt (EDTA, 99–101% purity) was obtained from Pancreac (Barcelona, Spain). The solvents for chromatography acetonitrile (99.9% purity) and methanol (≥99.9% purity) were acquired from ChemLab (Zedelgem, Belgium), and formic acid (98–100% purity) from Supelco-Sigma-Aldrich (Madrid, Spain). The standard L-Ascorbic acid (≥99.9% purity) was purchased from Sigma-Aldrich, while the standards Kuromanin chloride (cyanidin 3-*O*-glucoside chloride) (≥96% purity), Peonidin 3-*O*-glucoside chloride (≥95% purity), Oenin chloride (malvidin 3-*O*-glucoside chloride) (≥95% purity) and Myrtillin chloride (delphinidin 3-*O*-glucoside chloride) (≥95% purity) were provided by Extrasynthese (Genay, France).

2.2. Plant material

A total of 30 *Lactuca* accessions were included in this study (Table 1, Supplementary Fig. 1): 10 commercial lettuce varieties (4 green and 6 red), 13 traditional lettuce varieties (9 green and 4 semi-red), and 7 wild relative species. Among the lettuce varieties, 6 out of the 12 types defined by UPOV (2019), Buttherhead, Batavia, Frisée d'Amérique, Lollo, Cos and Gem, are represented. One of the wild species, closely related to *L. sativa*, is included in the primary gene pool (*L. dregeana* DC.), and the rest are more distant, forming part of the secondary or tertiary gene pool (PGR lettuce, www.pgrportal.nl/en/Lettuce-genetic-resources-Portal.htm), though this classification can vary depending on the methodology used to classify them (see Lebeda et al., 2004 and references herein included). In winter 2018/19, three plants per accession were grown in pots (30 × 25 cm and 11.7 L volume) with a mix of black and blonde peat (1:1) supplemented with fertilizer in a greenhouse at Agrifood Research and Technology Centre of Aragón (CITA, Zaragoza, Spain) following a completely randomized block design. No supplementary light was supplied so basal levels of phytochemicals (i.e. anthocyanins) could be measured.

After a period ranging from 2 and a half to 3 and a half months (depending on the accession), the plants were harvested. In all commercial and traditional lettuce varieties, as well as in the wild species with a rosette growth form (i.e. *Lactuca dentata* Makino, *Lactuca homblei* De Wild, and *Lactuca virosa* L.), inner and outer leaves were collected to represent the whole plant. In the case of small wild species (i.e. *Lactuca biennis* (Moench) Fernald and *Lactuca floridana* L. Gaertn), the whole plant was collected. Finally, in the two wild species with an early bushy growth (*L. dregeana* DC. and *Lactuca squarrosa* (Thunb.) Miq.), parts of the main stem were also sampled separate from the leaves. All samples were immediately frozen with liquid nitrogen and then kept at −80 °C until use.

2.3. Sample treatment

2.3.1. Extraction of vitamin C

Vitamin C extraction was carried out following the procedure described by Medina-Lozano et al. (2020). Briefly, 5 mL of extraction solution (8% acetic acid (v/v), 1% MPA (w/v) and 1 mM EDTA) were added to 50 mg of the lyophilized and finely powdered sample. The mixture was shaken in a vortex for 5 s and in an orbital shaker for 10 min at 2000 rpm. After being sonicated for 10 min at room temperature and centrifuged at 4000×*g* for 10 min at 4 °C, the supernatant was filtered through a 0.22-μm regenerated cellulose filter (Agilent, CA, United States) and recovered in a 5-mL amber vial. The obtained filtrate (Extract 1, E1) was used to determine vitamin C content in two steps. First, an aliquot of E1 was directly used to quantify AA as described below. Second, to obtain TAA, DHAA was reduced to AA by adding 200 μL of reducing solution (40 mM DTT with 0.5 M Tris pH 9.0) to a 200-μL aliquot of E1. After 30 min of reaction at room temperature in darkness, 200 μL of 0.4 M sulfuric acid were added to stop and stabilize AA in acidic pH (Extract 2, E2).

2.3.2. Extraction of anthocyanins

Anthocyanin extraction was performed in all the accessions, the green-leaf varieties included, according to the method described by Assefa et al. (2019) with slight modifications. Briefly, 5 mL of the extraction solution (methanol:ultrapure water:formic acid 50:44:6 v:v:v) was added to 40 mg of fine powder of the lyophilized sample. The mixture was vortexed for 5 s, shaken in an orbital shaker for 20 min at 2000 rpm and sonicated in an ultrasonic bath for 20 min at room temperature, followed by a centrifugation at 4000×*g* for 10 min at 4 °C. Supernatant was recovered in a 12-mL amber glass tube. A second extraction was carried out repeating the procedure just described. Both supernatants were blended and filtered through a 0.22-μm polytetrafluoroethylene (PTFE) filter (Agilent) to obtain the final extract (Extract 3, E3). The whole protocol was carried out under low intensity light.

2.4. UPLC conditions

All methods were conducted on the liquid chromatographic Acquity UPLC H-Class system equipped with an Acquity UPLC Photodiode Array PDA eλ Detector (Waters, Milford, MA, USA) and controlled by Empower 3 (Waters) software.

2.4.1. Chromatographic determination of vitamin C

Dilutions of E1 and E2 with ultrapure water (1:4 v:v) were prepared in 2-mL amber screw thread vials with caps and used to quantify AA and TAA, respectively, according to the method described by Medina-Lozano et al. (2020). An Acquity UPLC HSS T3 column (150 mm × 2.1 mm × 1.8 μm, Waters) was used and the mobile phases consisted of methanol (A) and ultrapure water pH 2.0 acidified with formic acid (C) with a flow rate of 0.3 mL min^{-1} of 2% A and 98% C in isocratic mode. The samples and column were kept at 5 °C and 30 °C, respectively. The injection volume was 5 μL and the total running time was 3 min. The wavelength of the detector was set at 245 nm. AA and TAA content were quantified using a calibration curve built with standard solutions at concentrations of 0.5, 2.5, 5.0, 10.0 and 25.0 μg mL^{-1} prepared with the stock solution

(100 μg mL^{-1}) of the AA standard using ultrapure water pH 2.0 acidified with formic acid.

2.4.2. Chromatographic determination of anthocyanins

Anthocyanin extract E3 was transferred to a 2-mL amber screw thread vial with a cap. The chromatographic procedure for anthocyanin determination was based in the method described by Fernández-Barbero et al. (2019) with some modifications. The separations were carried out in an Acquity UPLC BEH C18 column (150 mm × 2.1 mm × 1.7 μm, Waters). The mobile phases consisted of methanol (A) and ultrapure water pH 2.0 acidified with formic acid (C) with a flow rate of 0.3 mL min^{-1} in gradient mode of A and C (Supplementary Table 1). The injection volume was 3 μL and the total running time 20 min. The temperature of the samples and the column were adjusted to 10 °C and 30 °C, respectively, and the UV chromatograms were recorded at 520 nm.

Consistent patterns described in previous works in lettuce (Wu & Prior, 2005; Mulabagal et al, 2010; Becker et al., 2014) allowed us to identify two of the anthocyanins, cyanidin 3-*O*-(6′-*O*-malonylglucoside) and cyanidin 3-(6″-acetylglucoside), present in the samples included in this study. For their quantification, a calibration curve from 0.1 to 50 μg mL^{-1} of cyanidin 3-*O*-glucoside chloride was built. For peonidin 3-*O*-glucoside identification and quantification, a calibration curve from 0.01 to 1 μg mL^{-1} of peonidin 3-*O*-glucoside chloride was generated. No other anthocyanins were found.

2.5. Method validation

To achieve chromatographic separations, the methods were validated by determining different analytical parameters as described by Bertolín et al. (2018). First of all, the selectivity was analysed as the possibility of obtaining a signal free of interferences of other compounds for each analyte. The sensitivity was determined as the slope of the calibration curve. The limit of detection (LOD) and the limit of quantification (LOQ) were established as 3 and 10 times the standard deviation of 10 blanks (5 μL of mobile phase), respectively. The linear range was determined by using five increasing concentrations of the corresponding standards and linearity was presented as a coefficient of determination (R^2), considering an appropriate linearity if this coefficient reached at least a value of 0.970.

The repeatability and the intermediate precision were expressed as coefficients of variation (CV, %) and were determined as standard deviation/average × 100, being n = 5 (five samplings of the matrix, samples treatments and chromatographic conditions within the same day, in the case of repeatability, and on different days in the case of intermediate precision). Finally, the recovery (Rec, %) was calculated using 10 aliquots with 50 mg (AA) or 40 mg (anthocyanins) of the same sample, 5 spiked with 2 mg of analyte g^{-1} of dry matter, and 5 non spiked and was calculated as follows:

$$\frac{[\text{Analyte}]_{\text{spiked sample}} - [\text{Analyte}]_{\text{sample}}}{[\text{Analyte}]_{\text{spiked}}} \times 100$$

where n = 5. Furthermore, AA and anthocyanin stability were checked through consecutive injections of the same sample from 0 to 24 h (data not shown).

2.6. Statistical analysis

Summary statistics were calculated with the data coming from three biological replicates for each accession (n = 3). Effects of the group (according to Table 1), leaf colour and tissue on vitamin C content were tested by ANOVA with post hoc Tukey's test in the whole set of accessions ($p < 0.05$), the lettuce varieties and the two branching wild species (Supplementary Fig. 1), respectively. Regarding the total anthocyanin content, the effects of groups and tissues were checked as described above. Data were normalized when required. Pair-wise Spearman's correlation coefficient and significance ($p < 0.05$) was calculated between seven out of the eight traits measured in all the accessions (cyanidin 3-(6″-acetylglucoside) was only found in one sample, Table 2).

Hierarchical Clustering analysis was applied independently to the vitamin C and total anthocyanin data in all the accessions using the Ward's minimum variance to calculate the distance among clusters.

All statistical analyses were performed using JMP v5.1.2 software for Windows (SAS Institute Inc., Cary, NC).

3. Results and discussion

3.1. Method validation

All analytes showed an appropriate selectivity as they presented a signal free of interferences from other components of similar behaviour. The retention times for AA and anthocyanins were short, 1.874 min and 6.728 min, respectively, as well as the total running times, 3 min and 20 min, respectively.

The statistical parameters were optimized. All compounds presented a good linearity in the linear range tested (0.5–25 μg mL^{-1} for AA, 0.1–50 μg mL^{-1} for cyanidin 3-*O*-glucoside chloride, and 0.01–1 μg mL^{-1} for peonidin 3-*O*-glucoside chloride), with the following coefficients of determination: $R^2 = 0.99998$ for AA, $R^2 = 0.99831$ for cyanidin 3-*O*-glucoside chloride, and $R^2 = 0.99126$ for peonidin 3-*O*-glucoside chloride. The LODs were 13.05 μg AA g^{-1} DW (dry weight), 0.51 μg cyanidin 3-*O*-glucoside chloride g^{-1} DW, and 0.25 μg peonidin 3-*O*-glucoside chloride g^{-1} DW, while the LOQs were 45.21 μg AA g^{-1} DW, 1.77 μg cyanidin 3-*O*-glucoside chloride g^{-1} DW, and 0.87 μg peonidin 3-*O*-glucoside chloride g^{-1} DW.

The analytical methods presented good values of repeatability and intermediate precision: 1.75% and 4.22% for AA, 2.17% and 5.76% for cyanidin 3-*O*-glucoside chloride, and 3.19% and 5.91% for peonidin 3-*O*-glucoside chloride, respectively. Finally, the recovery was also good for AA, cyanidin 3-*O*-glucoside chloride, and peonidin 3-*O*-glucoside chloride: 95.6%, 97.2%, and 96.3%, respectively.

Furthermore, AA and TAA stability was analysed as AA is a very labile molecule that oxidizes easily at high temperatures, high pH, under intense light and in oxidizing atmosphere (Lee and Kader, 2000). AA and TAA degradations started 4 h after extraction (data not shown). Hence, it was necessary to quantify them in this time interval. A short total running time was needed to analyse a high number of samples in the 4 h window. In the case of anthocyanins, their stability was tracked over a 24 h period, without observing significant losses in the analyte concentrations. Chromatograms of 3 and 20 min were achieved without interferences between two consecutive samples in the case of AA and anthocyanins, respectively.

3.2. Quantification of vitamin C

The mean TAA content ranged between 153.24 and 291.11 mg 100 g^{-1} DW in leaf tissue (Table 2, Fig. 1a), the poorest accession being a red-leaf commercial variety ('Nestorix'), while the richest was a wild relative (*L. homblei*). In general, the wild species had higher content of both, DHAA and AA, and consequently also of TAA. In traditional lettuces, the average content of these three compounds was also higher than in the commercial varieties. These results are in agreement with those obtained by van Treuren et al. (2018), that also observed the highest content of AA in primitive lettuces and their wild relatives. This supports the idea that wild *Lactuca* species could play an important role in modern lettuce breeding. Until now, lettuce wild relatives have been mainly used to introduce resistance genes in modern varieties but these results make evident their great potential to improve the nutritional value of a vegetable which is relatively poor in macronutrients like vitamin C (USDA, 2016). When all raw data were taken into account, the plant with the highest content in total vitamin C happened to belong to a

Table 2
Average weight, vitamin C and anthocyanin content (n = 3) in commercial and traditional lettuce varieties and some wild relatives (*Lactuca* spp.).

			Vitamin C (mg 100 g^{-1} DW)			Anthocyanins (mg 100 g^{-1} DW)[a]			
Sample	Group	Weight (g)	DHAA	AA	TAA	Peonidin 3-*O*-glucoside	Cyanidin 3-*O*-(6′-*O*-malonylglucoside)	Cyanidin 3-(6″-acetylglucoside)	Total
'Begoña'	Commercial varieties (green)	102.43 ± 31.64	28.55 ± 3.09	255.48 ± 21.82	284.03 ± 20.65	ND	ND	ND	ND
'Dolomiti G12′		115.17 ± 7.77	26.67 ± 9.49	172.67 ± 9.43	199.35 ± 6.99	ND	ND	ND	ND
'Romana Inverna'		238.63 ± 13.74	19.98 ± 1.49	228.51 ± 24.87	248.50 ± 23.57	ND	ND	ND	ND
'Winter Crop'		208.93 ± 47.73	44.97 ± 15.21	118.74 ± 20.63	163.71 ± 20.63	ND	ND	ND	ND
'Likarix'	Commercial varieties (red)	297.97 ± 29.58	14.43 ± 3.60	205.29 ± 8.27	219.71 ± 5.24	3.37 ± 1.74	123.91 ± 20.50	ND	127.28 ± 21.64
'Lollo Rosso'		291.70 ± 17.68	25.36 ± 5.26	147.09 ± 19.69	172.45 ± 19.70	ND	8.95 ± 1.56	ND	8.95 ± 1.56
'Nestorix'		329.13 ± 16.60	22.12 ± 4.50	131.12 ± 37.94	153.24 ± 40.94	2.60 ± 1.44	55.41 ± 15.13	ND	58.01 ± 16.56
'Red Sails'		343.67 ± 34.70	30.51 ± 5.16	167.01 ± 22.06	197.52 ± 27.14	1.78 ± 0.48	59.58 ± 18.46	ND	61.36 ± 18.73
'Revolution'		215.67 ± 28.76	31.16 ± 11.02	157.32 ± 9.83	188.48 ± 9.70	0.95 ± 0.76	37.91 ± 22.20	ND	38.86 ± 22.96
'Romired'		318.23 ± 32.80	45.79 ± 6.53	150.22 ± 23.66	196.02 ± 17.13	2.34 ± 1.02	58.06 ± 15.99	ND	60.40 ± 16.73
'Lechuga de Beceite'	Traditional varieties (green)	208.80 ± 60.60	21.53 ± 8.35	213.46 ± 72.65	234.99 ± 71.88	ND	ND	ND	ND
'Lechuga de Ensalada'		241.33 ± 47.82	45.53 ± 8.90	164.46 ± 34.14	210.00 ± 40.02	ND	ND	ND	ND
'Lechuga de Híjar'		343.57 ± 83.22	22.94 ± 3.77	165.19 ± 52.27	188.13 ± 54.11	ND	ND	ND	ND
'Lechuga de Subías'		226.30 ± 58.00	21.01 ± 10.14	204.97 ± 26.44	225.98 ± 32.86	ND	ND	ND	ND
'Lechuga del Pirineo'		244.17 ± 109.82	25.45 ± 15.73	238.63 ± 84.74	264.08 ± 100.28	ND	ND	ND	ND
'Lechuga del Valle de Tena'		295.63 ± 45.31	47.03 ± 7.87	165.52 ± 26.94	212.54 ± 26.88	ND	ND	ND	ND
'Lechuga Romana Zaragozana'		163.40 ± 45.80	10.43 ± 7.39	270.38 ± 22.74	280.81 ± 30.13	ND	ND	ND	ND
'Lengua de Buey'[b]		188.97 ± 80.02	31.85 ± 7.97	178.05 ± 47.04	209.91 ± 45.99	ND	1.49 ± 2.59	ND	1.49 ± 2.59
'Oreja de Mulo'		219.90 ± 101.99	29.92 ± 15.64	218.75 ± 60.35	248.68 ± 71.16	ND	ND	ND	ND
'Lechuga de Bureta'	Traditional varieties (semi-red)	171.70 ± 69.04	24.40 ± 3.74	232.97 ± 25.70	257.37 ± 22.20	0.44 ± 0.77	22.48 ± 33.96	ND	22.93 ± 34.72
'Morada de Belchite'		298.57 ± 50.93	10.25 ± 7.63	147.88 ± 30.92	158.13 ± 23.99	0.35 ± 0.61	11.99 ± 14.51	ND	12.34 ± 15.11
'Morada de Bernués'		252.77 ± 37.04	14.61 ± 0.84	183.32 ± 18.39	197.93 ± 18.75	ND	0.50 ± 0.30	ND	0.50 ± 0.30
'Morada de Sorripas'		297.60 ± 18.30	24.07 ± 4.61	157.53 ± 18.82	181.60 ± 14.42	ND	11.41 ± 1.72	ND	11.41 ± 1.72
L. biennis	Wild species (leaf)	4.57 ± 1.56	53.09 ± 11.13	164.68 ± 11.56	217.77 ± 9.51	ND	1.05 ± 0.86	ND	1.05 ± 0.86
L. dentata		38.63 ± 17.33	23.30 ± 4.93	199.05 ± 15.61	222.35 ± 14.21	ND	16.38 ± 5.36	ND	16.38 ± 5.36
L. dregeana		221.90 ± 25.34	40.03 ± 14.85	225.79 ± 16.10	265.82 ± 28.90	ND	12.51 ± 6.34	ND	12.51 ± 6.34
L. floridana		1.23 ± 0.29	150.05 ± 5.87	122.87 ± 18.62	272.92 ± 12.75	ND	14.13 ± 1.09	ND	14.13 ± 1.09
L. homblei		116.27 ± 3.00	50.27 ± 5.41	240.84 ± 34.35	291.11 ± 38.26	ND	3.56 ± 0.46	ND	3.56 ± 0.46
L. squarrosa		186.50 ± 68.64	53.40 ± 22.16	233.13 ± 69.95	286.53 ± 48.76	ND	2.26 ± 1.28	ND	2.26 ± 1.28
L. virosa		102.27 ± 29.71	40.91 ± 6.04	188.45 ± 5.91	229.36 ± 4.61	ND	0.98 ± 0.55	ND	0.98 ± 0.55
L. dregeana	Wild species (stem)	–	17.11 ± 3.08	81.26 ± 25.33	98.38 ± 28.23	ND	15.39 ± 4.34	ND	15.39 ± 4.34
L. squarrosa		–	13.23 ± 5.33	34.27 ± 19.30	47.50 ± 24.25	ND	17.53 ± 15.18	1.95 ± 0.28	19.48 ± 15.23

[a]ND: not detected.
[b]Only one out of the three biological repeats contained anthocyanins.

traditional variety ('Lechuga del Pirineo'), which is a romaine-type lettuce with good organoleptic and agronomic characteristics (Carravedo et al., 2011).

In general, the amounts of vitamin C obtained are higher than those reported in other studies carried out in the same varieties (i.e. 'Lollo Rosso' in Llorach et al., 2008) or in wild *Lactuca* species (i.e. *L. dentata* and *L. virosa* in van Treuren et al., 2018). This is most certainly due to the fact that the starting materials employed are lyophilized leaves, whereas most authors have conducted their experiments using fresh leaves, in which all the compounds are expected to be more diluted. In agreement with this, the results are in the same order of magnitude (even if they are generally higher) than those reported by Złotek et al. (2014), who used lyophilized leaves of *L. sativa* L. *var. capitata*, a variety not included in our study.

Except for *L. floridana*, where DHAA was the most abundant form of vitamin C, AA was the main contributor to the total vitamin C content in all accessions (Table 2, Fig. 1a), representing on average 85% of the TAA. This agrees with observations in most vegetables, where AA is also the most abundant form of vitamin C (Lee and Kader, 2000). However, even if AA is also the form showing the highest biological activity (antioxidant), DHAA is easily converted into AA in the human body (Lee & Kader, 2000). So, DHAA can serve as vitamin C reservoir in certain adverse conditions (i.e. oxidative stress, which is a consequence derived from other stresses, like drought). In this sense, genetic resources like lettuce wild relatives are awaking interest among breeders to increase the crop tolerance to an array of abiotic stresses (Hartman et al., 2014).

When contents of DHAA, AA and TAA were compared among the three groups studied, differences in their average values were observed (Table 2, Fig. 2a), which became significant in the case of TAA ($F = 4.438$, $p = 0.022$), and very significant for DHAA ($F = 6.786$, $p = 0.004$). The wild relatives showed up to 21% and 51% more TAA and DHAA, respectively, when compared to the commercial varieties. In the case of the traditional varieties, the average content of DHAA was similar to the amount present in the commercial varieties, whereas the TAA value was 8% higher. Regarding AA, differences could be also observed, even if they did not reach statistical significance. In fact, the groups formed by the lettuce wild relatives and the traditional varieties exhibited 12% and 11% more AA, respectively, than the lettuce commercial varieties. Until now, lettuce breeding has been focused mainly on obtaining crops that are more productive and resistant to biotic stresses, however little attention has been paid to its nutritional quality. This could have led to a phytonutrient "wash" in modern varieties, which has also been observed in other crops, like apple, where the selection of sweeter and less acidic fruit has resulted in a decrease in their content in vitamin C when wild apples are compared to cultivated ones (Fang et al., 2017). This would explain why non-domesticated plants (Crop Wild Relatives, CWR) or those under low selection pressure (traditional varieties), show higher levels, in this case, of vitamin C.

Considering only the cultivated lettuces, both commercial and traditional varieties, the results obtained allowed to conclude that green-leaf varieties contained higher amounts of vitamin C in all its forms (DHAA, AA and TAA) than red-leaf varieties (Fig. 2b), with up to a 15% more in the case of DHAA, and 18% more in AA and TAA content. Specifically, the differences were significant ($F = 5.480$, $p = 0.029$) and very significant ($F = 8.968$, $p = 0.007$) in AA and TAA content, respectively, between green and red varieties. No interaction between group and leaf colour was detected (data not shown). No exhaustive work comparing both types of lettuce in terms of vitamin C content is available in the literature. However, Llorach et al. (2008) analysed four varieties, two green and two red, and found that the two showing both the highest and the lowest vitamin C contents, were the two green varieties.

The two wild species with an early bushy growth (*L. dregeana* DC. and *L. squarrosa*, Supplementary Fig. 1) offered the opportunity to collect samples from two different tissues, leaf and stem, at the moment of harvest. Both species accumulated more DHAA, AA and TAA in the leaves than in the stems (Table 2, Fig. 2c). In all cases, the differences observed were significant ($F = 20.541$ and $p = 0.045$; $F = 52.119$ and $p = 0.019$; $F = 54.755$ and $p = 0.018$ for DHAA, AA and TAA, respectively).

3.3. Quantification of anthocyanins

Without considering the green-leaf varieties, in which in most cases there was a complete absence of anthocyanins, the average values of total anthocyanins ranged between 0.50 and 127.28 mg 100 g^{-1} DW in leaf tissue (Table 2 and Fig. 1b). The accession with the lowest content was a semi-red traditional variety ('Morada de Bernués') and the one with the highest value was a red commercial variety ('Likarix'). Surprisingly, all wild species assayed contained anthocyanins, almost exclusively cyanidin 3-*O*-(6′-*O*-malonylglucoside), even those apparently being green-leaf coloured, like *L. dentata* (Supplementary Fig. 1). This same anthocyanin was also present in the green-leaf traditional variety 'Lengua de Buey' (Table 1, Supplementary Fig. 1), even if it had been previously described as a lettuce with yellow-green leaves and no anthocyanic colour (Carravedo et al., 2011). Trace amounts of anthocyanins in green-leaf varieties have been previously reported (Kleinhenz et al., 2003; Brücková et al., 2016). However, the cyanidin 3-*O*-(6′-*O*-malonylglucoside) was only found in one out of the three plants analysed from this accession ('Lengua de Buey'-2). These differences (presence/absence of a particular anthocyanin) within the same accession, show the enormous heterogeneity and diversity harboured by the traditional varieties. Cyanidin 3-*O*-(6′-*O*-malonylglucoside) is not only the most abundant anthocyanin, as found before (García-Macías et al., 2007), representing the 97% of the total anthocyanin content on average, but also the only one present in most samples (Fig. 1b). Furthermore, it appears across all the groups, commercial and traditional lettuce varieties, as well as all the wild species studied here. Quite the opposite, peonidin 3-*O*-glucoside is only present in cultivated forms (*L. sativa*). Besides, it is always detected simultaneously with cyanidin 3-*O*-(6′-*O*-malonylglucoside), although in much lower amounts. The cyanidin glucosides are transformed into the peonidin glucosides in a single step catalysed by a methyltransferase. That is why they are commonly found together. Interestingly, the cyanidin 3-(6″-acetylglucoside) was present only in one of the wild species (*L. squarrosa*) and, within it, exclusively in the stem and not in the leaf (Table 2, Fig. 1b). This is the first time that tissue-dependent biosynthesis of a specific anthocyanin has been reported in *Lactuca*, though anthocyanin presence in only certain tissues has been described in *Citrus* genus (Fabroni et al., 2016).

Cyanidin-type is also the most abundant among the anthocyanins present in many other crops, like mulberry (Kim & Lee, 2020), purple corn (Peniche-Pavía & Tiessen, 2020), purple wheat (Abdel-Aal et al., 2018), blood orange (Fabroni et al., 2016), with glucose being the main sugar and malonyl the predominant acyl substituent in many cases, lettuce included (Wu and Prior, 2005; Mulabagal et al., 2010; Becker et al., 2014), as observed in this work (Table 2, Fig. 1b). Regarding the other two minor anthocyanins, cyanidin 3-(6″-acetylglucoside), which was only found in *L. squarrosa* stem, and peonidin 3-*O*-glucoside, both have been rarely reported in lettuce before (Wu and Prior, 2005; Baek et al., 2013; Viacava et al., 2017).

In contrast to what was observed for vitamin C, on average, commercial lettuce varieties had higher content of total anthocyanins than traditional lettuce varieties, and the latter, higher than the wild species (Fig. 2d). However, only the differences between the commercial varieties and the other two groups became very significant statistically ($F = 8.452$, $p = 0.003$). So, this ranking shows the inverse order than the one observed for the average vitamin C content. This effect can be attributed to the fact that the red colour has been enhanced in the new red varieties, as pointed out before (Casals Missio et al., 2018), because it makes the product more attractive for consumers and increases its potential market value (not to mention the benefits for human health derived from a higher anthocyanin content). In this study, commercial varieties

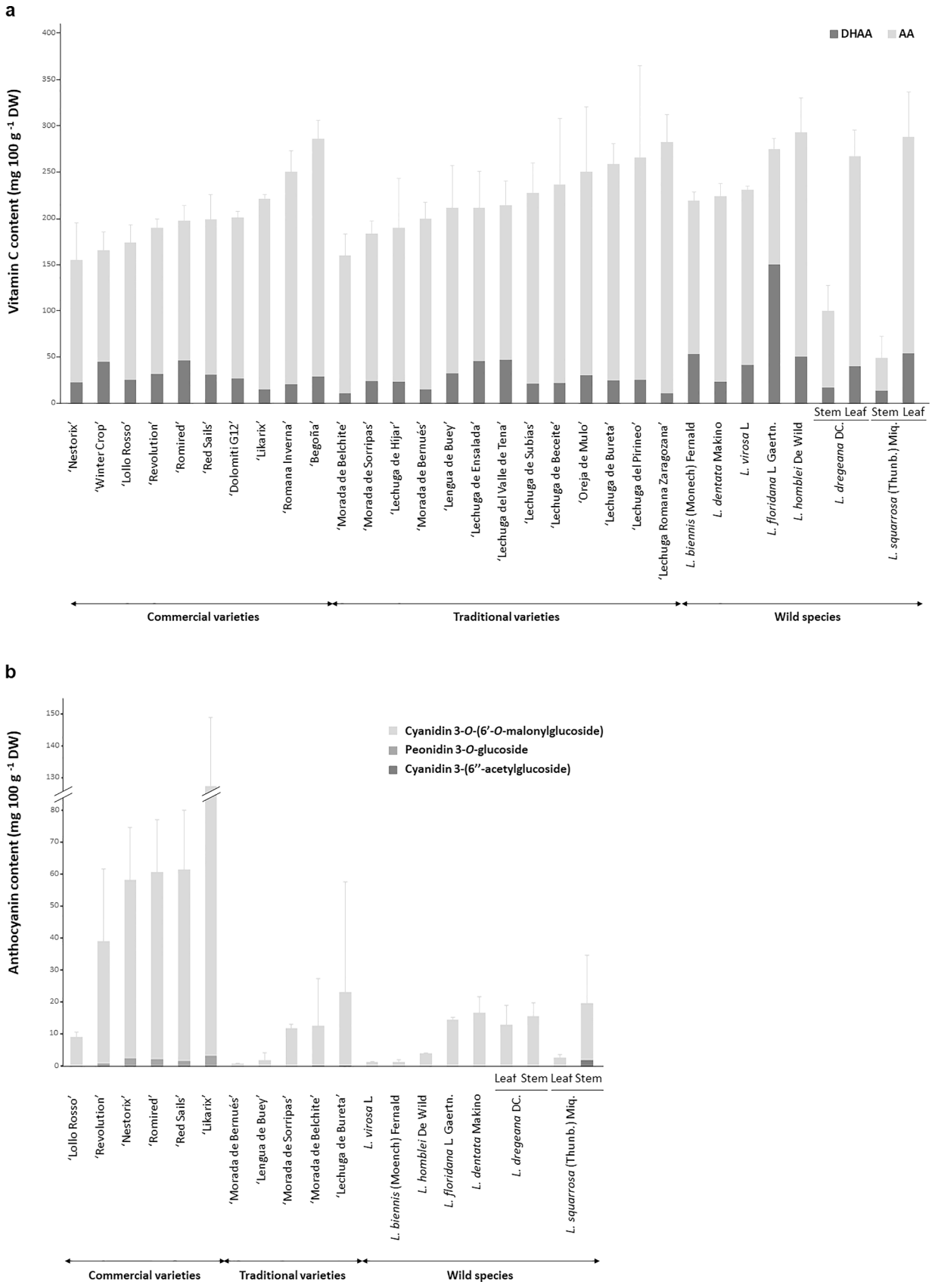


Fig. 1. Total concentration of vitamin C (a) and anthocyanins (b) present in lettuce varieties (commercial and traditional) and some wild relatives. Bars represent the standard deviation of the total (n = 3).

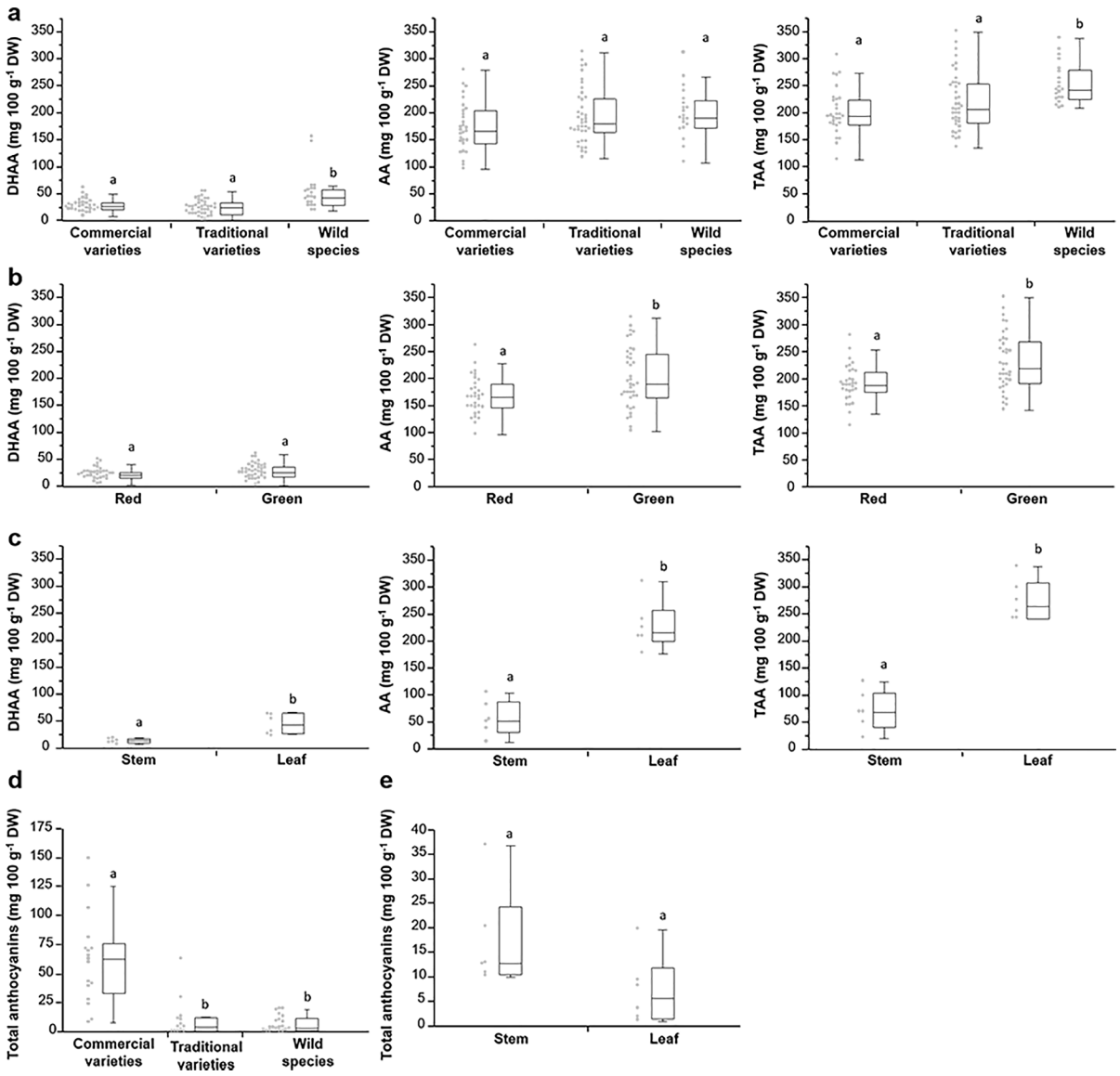


Fig. 2. Box plots of nutrient content: DHAA, AA and TAA in commercial and traditional lettuce varieties and some wild relatives (a), in green and red-leafed lettuce varieties (b), and in stem and leaf of two wild species (see Table 1) (c); total anthocyanins in commercial and traditional lettuce varieties and some wild relatives (d), and in stem and leaf of two wild species (see Table 1) (e). Different letters indicate significant differences ($p < 0.05$) among groups. Note: 'Lengua de Buey'-1, -3 included in the green group and -2 in the red group.

accumulated six and eight times more anthocyanins than traditional varieties and the wild relatives analysed, respectively. However, there are some exceptions, like *L. dentata*, *L. dregeana* and *L. floridana*, in which the mean of the total anthocyanins was higher than in a commercial variety like 'Lollo Rosso' (Table 2), under the particular growing conditions described above. A similar trend was observed in three out of the four semi-red traditional varieties analysed, 'Lechuga de Bureta', 'Morada de Belchite' and 'Morada de Sorripas', which also accumulated more anthocyanins on average than 'Lollo Rosso' (Table 2, Fig. 2b).

A direct comparison of anthocyanin accumulation with previous works is not possible as, in most cases, it has been studied only in one or a few lettuce varieties and under different growing conditions (temperature, illumination, fortification with micronutrients, greenhouse *vs* outdoor cultivation, etc). However, evaluation of lettuce germplasm for anthocyanin content, including cultivated forms (commercial and traditional) as well as wild relatives, has not been carried out before. Even comparing a particular accession, for instance 'Lollo Rosso', the anthocyanin level obtained is both, higher (Ordidge et al., 2010) and lower (García-Macías et al., 2007), than those found under other control conditions in other researches (taking also into account that we referred it to DW unlike in those studies). This is not surprising as anthocyanin synthesis is hugely affected by a number of factors (Chalker-Scott, 1999).

Unlike the vitamin C content in the tissues coming from the wild species *L. dregeana* and *L. squarrosa*, the anthocyanin content was higher in the stem than in the leaf (Table 2), however no significant differences were observed (Fig. 2e). As commented before, the cyanidin 3-(6″-acetylglucoside) was found only in the stem of *L. squarrosa*, but not in the leaves. Furthermore, in this case, this cyanidin glucoside was present in the three biological replicates of *L. squarrosa* tested. Cyanidin 3-(6″-acetylglucoside) has not been identified in any of the cultivated lettuces included in this study, though it has been found previously in some red lettuce varieties (Wu and Prior, 2005; Viacava et al., 2017).

3.4. Pairwise correlations between traits

As expected, AA showed a negative and a positive correlation with DHAA and TAA, respectively, though only the correlation between AA

and TAA was strong and statistically significant (Spearman's rho = 0.868, $p < 0.0001$, Table 3). Those coefficients reflect that AA and DHAA are interconvertible (one appears at the expense of the other) and also that the main contribution to TAA is made by AA. Interestingly, the correlation between total anthocyanins, cyanidin 3-*O*-(6′-*O*-malonylglucoside) and peonidin 3-*O*-glucoside with both, AA and TAA, are all negative. Even if the values of the coefficients were low, in some cases, they were statistically significant (Table 3). Previously, it has been suggested that the biosynthesis of anthocyanins could cause a reduction in other processes of the plant as it is a direct competition for assimilated carbon (García-Macías et al., 2007). This could explain why the red-leaf varieties contained a significantly lower amount of TAA than the green-leaf varieties (Fig. 2b). This also agrees with the inverse ranking observed for average TAA (commercial varieties < traditional varieties < wild relatives) when compared to average total anthocyanins (wild relatives < traditional varieties < commercial varieties). The two main anthocyanins found, cyanidin 3-*O*-(6′-*O*-malonylglucoside) and peonidin 3-*O*-glucoside, show a positive and highly significant correlation (Spearman's rho = 0.717, $p < 0.0001$). This coincides with their simultaneous presence in most of the lettuce varieties studied (8 out of 10, Table 2) and with the peonidin synthesis simply by transferring a methyl group to cyanidin's B ring at 3′ position (Zhang et al, 2014). Both of them were also positively correlated at a highly significant level with the total anthocyanin content, especially the most abundant (cyanidin 3-*O*-(6′-*O*-malonylglucoside), which showed a perfect positive correlation (Spearman's rho = 1, $p < 0.0001$).

Finally, the plant weight showed a negative relationship with the vitamin C content (DHAA, AA and TAA), weak though significant in the case of DHAA ($p = 0.042$) and highly significant for AA and TAA ($p < 0.0001$); whereas it was positive with the two main anthocyanins, as well as with the total anthocyanin content, being highly significant with the peonidin 3-*O*-glucoside ($p < 0.0001$) (Table 3). A negative correlation between biomass and vitamin C content has been previously described in lettuce (Bumgarner et al., 2012) and could be explained by the fact that to synthesize it results costly for the plant cell. The same could be argued for the anthocyanins and, actually, a negative correlation between the biomass and total anthocyanin content has been reported by the same authors (Bumgarner et al., 2012) though, under the experimental conditions of our study, none of the accessions developed a strong anthocyanic colour (Supplementary Fig. 1).

3.5. Nutritional value

The nutritional value of the samples analysed here, expressed as vitamin C and total anthocyanin content, respectively, showed great variation, especially in the case of anthocyanins (Fig. 1). Their hierarchical clustering allowed us to create groups of accessions with similar content of each of the two compounds separately (Fig. 3). Regarding the vitamin C (Fig. 3a), the accessions organised themselves in two main groups (distance = 2.16): high (248–287 mg 100 g^{-1} DW, Cluster I) and medium–low (153–235 mg 100 g^{-1} DW, Cluster II) content. Within Cluster II, subgroups with the lowest (153–198 mg 100 g^{-1} DW, Subcluster IIa), and medium (210–235 mg 100 g^{-1} DW, Subcluster IIb) amount of vitamin C could be distinguished. It is noteworthy that the richest (Cluster I) and the intermediate (Subcluster IIb) groups in vitamin C are mainly formed by traditional varieties and lettuce wild relatives. These results could promote the ongoing initiative to valorise traditional crop varieties and landraces as a source of highly nutritious compounds (Petropoulos et al., 2019). In this sense, Casals Missio et al. (2018) found that lettuce landraces were characterised by higher levels of sugars when compared to modern varieties. In the other extreme, 7 out of the 11 accessions included in the poorest group in terms of vitamin C concentration (Subcluster IIa) are commercial varieties, mainly read-leaf lettuces.

In relation to the total anthocyanins (Fig. 3b), the accessions arranged themselves in two major groups (distance = 2.30): nothing to moderate (0–22.93 mg 100 g^{-1} DW, Cluster I) and high (38.86–127.28 mg 100 g^{-1} DW, Cluster II) amount. In Cluster I, subgroups with nothing or a negligible (0–3.56 mg 100 g^{-1} DW, Subcluster Ia) and medium (11.41–22.93 mg 100 g^{-1} DW, Subcluster Ib) amount of anthocyanins were obtained. Obviously, the poorest group (Subcluster Ia) was mainly integrated by green-leaf lettuces though there was also a semi-red traditional variety and four lettuce wild relatives. The remaining semi-red traditional varieties fall into the medium group (Subcluster Ib) together with the rest of the wild species. The richest group in total anthocyanins (Cluster II) was exclusively formed by red-leaf commercial varieties, with 'Likarix' exceling as it contained one order of magnitude more anthocyanins than the rest. By contrast, 'Lollo Rosso' did not develop a strong red colour (Supplementary Fig. 1) under our specific growing conditions, behaving like the semi-red traditional varieties in relation to the anthocyanin biosynthesis. As commented before, the red-leaf commercial varieties have been bred to accumulate higher concentrations of anthocyanins; this became clear when they were compared to red-leaf landraces performing a colour analysis (Casals Missio et al., 2018).

4. Conclusions

The current market trend reflects the growing interest of consumers in food quality. This justifies making an effort in improving the nutritional value of food, particularly, vegetables. For that, the first logical step should be to evaluate the content of health-promoting compounds in the main crops and their related germplasm. In this sense, wild crop relatives as well as landraces and traditional varieties play a pivotal role, especially the last two as they represent a shortcut in the long path to obtain new biofortified varieties. The characterisation of vitamin C and anthocyanin content of a wide variety of cultivated (commercial and traditional varieties) lettuces and wild forms, has been carried out. Lettuce wild relatives and commercial varieties were the richest groups in vitamin C and anthocyanins, respectively. Conversely, commercial varieties and wild relatives were the poorest groups in vitamin C and anthocyanins, respectively, with the traditional varieties occupying an

Table 3
Spearman's correlation coefficients and significance level ($p < 0.05$) between the traits studied (weight and vitamin C and anthocyanin concentrations) in commercial and traditional lettuce varieties, as well as wild relatives (*Lactuca* spp.).

	DHAA	AA	TAA	Peonidin 3-*O*-glucoside	Cyanidin 3-*O*-(6′-*O*-malonylglucoside)	Total anthocyanins	Weight
DHAA	1.000						
AA	−0.143	1.000					
TAA	0.243*	0.868***	1.000				
Peonidin 3-*O*-glucoside	−0.124	−0.209	−0.271*	1.000			
Cyanidin 3-*O*-(6′-*O*-malonylglucoside)	0.001	−0.217*	−0.180	0.717***	1.000		
Total anthocyanins	0.000	−0.218*	−0.181	0.717***	1.000***	1.000	
Weight	−0.220*	−0.473***	−0.538***	0.400***	0.209	0.210	1.000

***$p < 0.001$; *$p < 0.05$.

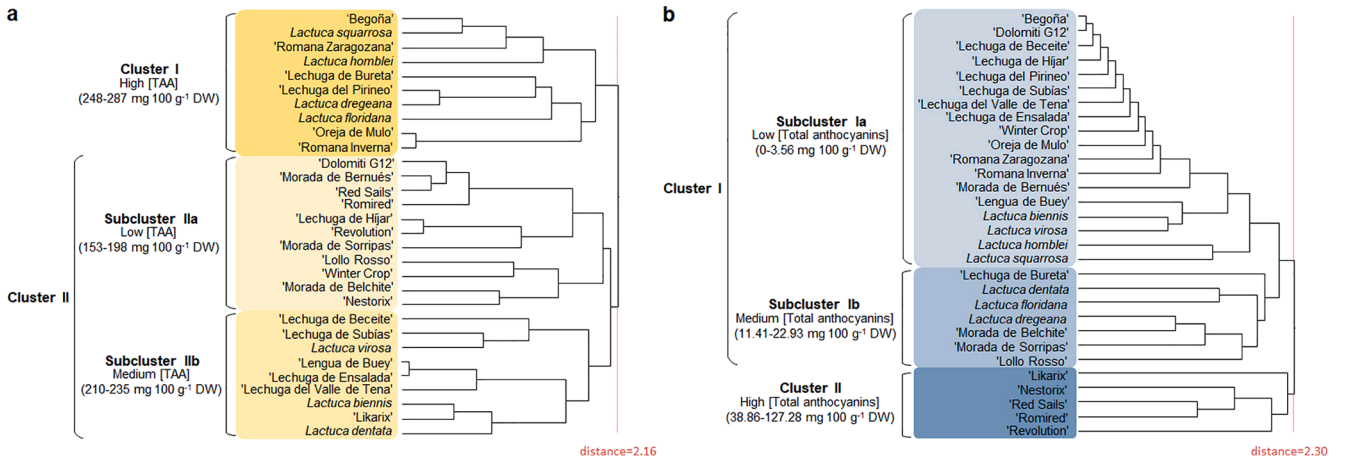


Fig. 3. Hierarchical clustering analysis of the 30 accessions included in this study (10 lettuce commercial varieties, 13 lettuce traditional varieties and 7 wild relatives) attending to their average vitamin C (a) and total anthocyanin (b) content (n = 3).

intermediate position in both cases. Some of the traditional lettuce varieties studied could be considered promising as they are rich in vitamin C and, in some cases, are able to biosynthesize anthocyanins, whose content could be increased by breeding.

The results offer nutritional information on what is the most consumed (but not especially nutritious) leafy vegetable and wild relatives, some of them with potential culinary use. This could be useful for both, consumers and breeders interested in improving the lettuce content of phytochemicals that promote health and reduce the risk of certain diseases.

CRediT authorship contribution statement

Inés Medina-Lozano: Experimentation, Analysis, Writing - review & editing. **Juan Ramón Bertolín:** Experimentation, Data collection, Analysis, Writing - review & editing. **Aurora Díaz:** Funding acquisition, Conceptualisation and design, Supervision, Project administration, Data collection, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank J. A. Aranjuelo, A. Castellanos and "laboratorio de valoración nutritiva" from CITA for technical support and D. L. Goodchild for reviewing the English language. This work was funded by the projects RTA2017-00093-00-00 from the National Institute *for* Agricultural and Food Research and Technology (INIA) and LMP164_18 from the Government of Aragón; and by the Operational Programme FEDER Aragón 2014-2020 and the European Social Fund from the European Union [Grupos Consolidados A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética" and A14-17R: "Sistemas agroganaderos alimentarios sostenibles" (SAGAS)]. I. M.-L. was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Spanish State Research Agency (AEI).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.129864.

References

Abdel-Aal, E. S. M., Hucl, P., & Rabalski, I. (2018). Compositional and antioxidant properties of anthocyanin-rich products prepared from purple wheat. *Food Chemistry, 254*, 13–19. https://doi.org/10.1016/j.foodchem.2018.01.170.

Assefa, A. D., Choi, S., Lee, J.-E., Sung, J.-S., Hur, O.-S., Ro, N.-Y., … Rhee, J.-H. (2019). Identification and quantification of selected metabolites in differently pigmented leaves of lettuce (*Lactuca sativa* L.) cultivars harvested at mature and bolting stages. *BMC Chemistry, 13*(1), 1–15. https://doi.org/10.1186/s13065-019-0570-2.

Baek, G. Y., Kim, M. H., Kim, C. H., Choi, E. G., Jin, B. O., Son, J. E., & Kim, H. T. (2013). The effect of LED light combination on the anthocyanin expression of lettuce. *IFAC Proceedings Volumes, 46*(4), 120–123. https://doi.org/10.3182/20130327-3-jp-3017.00029.

Becker, C., Klaering, H. P., Schreiner, M., Kroh, L. W., & Krumbein, A. (2014). Unlike quercetin glycosides, cyanidin glycoside in red leaf lettuce responds more sensitively to increasing low radiation intensity before than after head formation has started. *Journal of Agricultural and Food Chemistry, 62*(29), 6911–6917. https://doi.org/10.1021/jf404782n.

Bertolín, J. R., Joy, M., Rufino-Moya, P. J., Lobón, S., & Blanco, M. (2018). Simultaneous determination of carotenoids, tocopherols, retinol and cholesterol in ovine lyophilised samples of milk, meat, and liver and in unprocessed/raw samples of fat. *Food Chemistry, 257*, 182–188. https://doi.org/10.1016/j.foodchem.2018.02.139.

Brücková, K., Sytar, O., Živčák, M., Brestič, M., & Lebeda, A. (2016). The effect of growth conditions on flavonols and anthocyanins accumulation in green and red lettuce. *Journal of Central European Agriculture, 17*(4), 986–997. https://doi.org/10.5513/JCEA01/17.4.1802.

Bu, C., Zhang, Q., Zeng, J., Cao, X., Hao, Z., Qiao, D., … Xu, H. (2020). Identification of a novel anthocyanin synthesis pathway in the fungus *Aspergillus sydowii* H-1. *BMC Genomics, 21*(1), 1–16. https://doi.org/10.1186/s12864-019-6442-2.

Bumgarner, N. R., Scheerens, J. C., Mullen, R. W., Bennett, M. A., Ling, P. P., & Kleinhenz, M. D. (2012). Root-zone temperature and nitrogen affect the yield and secondary metabolite concentration of fall- and spring-grown, high-density leaf lettuce. *Journal of the Science of Food and Agriculture, 92*(1), 116–124. https://doi.org/10.1002/jsfa.4549.

Carr, A. C., & Frei, B. (1999). Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. *American Journal of Clinical Nutrition, 69*(6), 1086–1107. https://doi.org/10.1093/ajcn/69.6.1086.

Carravedo, M., Mayor, C., & Garcés-Claver, A. (2011) Evaluación morfológica y molecular de variedades autóctonas aragonesas de lechuga (Lactuca sativa L.) y especies silvestres emparentadas (Lactuca spp.): conservadas en el Banco de Germoplasma de Especies Hortícolas de Zaragoza. Descriptiva de las variedades de origen aragonés seleccionadas (pp. 107-170). Zaragoza: Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA)-Gobierno de Aragón.

Casals Missio, J., Rivera, A., Figàs, M. R., Casanova, C., Camí, B., Soler, S., & Simó, J. (2018). A comparison of landraces vs. modern varieties of lettuce in organic farming during the winter in the mediterranean area: An approach considering the viewpoints of breeders, consumers, and farmers. *Frontiers Plant Science, 9*(1491), 1–15. https://doi.org/10.3389/fpls.2018.01491.

Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology, 70*(1), 1–9. https://doi.org/10.1111/j.1751-1097.1999.tb01944.x.

Fabroni, S., Ballistreri, G., Amenta, M., & Rapisarda, P. (2016). Anthocyanins in different *Citrus* species: An UHPLC-PDA-ESI/MS^n-assisted qualitative and quantitative investigation. *Journal of the Science of Food and Agriculture, 96*(14), 4797–4808. https://doi.org/10.1002/jsfa.7916.

Fang, T., Zhen, Q., Liao, L., Owiti, A., Zhao, L., Korban, S. S., & Han, Y. (2017). Variation of ascorbic acid concentration in fruits of cultivated and wild apples. *Food Chemistry, 225*, 132–137. https://doi.org/10.1016/j.foodchem.2017.01.014.

FAOSTAT. Statistics of the Food and Agriculture Organization of the United Nations. (2018). http://www.fao.org/faostat/en/#data/QC/ Accesed 21 June 2020.

Fernández-Barbero, G., Pinedo, C., Espada-Bellido, E., Ferreiro-González, M., Carrera, C., Palma, M., & García-Barroso, C. (2019). Optimization of ultrasound-assisted extraction of bioactive compounds from jabuticaba (*Myrciaria cauliflora*) fruit through a Box-Behnken experimental design. *Food Science and Technology, 39*(4), 1018–1029. https://doi.org/10.1590/fst.16918.

García-Macías, P., Ordidge, M., Vysini, E., Waroonphan, S., Battey, N. H., Gordon, M. H., … Wagstaffe, A. (2007). Changes in the flavonoid and phenolic acid contents and antioxidant activity of red leaf lettuce (Lollo Rosso) due to cultivation under plastic films varying in ultraviolet transparency. *Journal of Agricultural and Food Chemistry, 55*(25), 10168–10172. https://doi.org/10.1021/jf071570m.

Hartman, Y., Hooftman, D. A. P., Uwimana, B., Schranz, M. E., Van de Wiel, C. C. M., Smulders, M. J. M., … Van Tienderen, P. H. (2014). Abiotic stress QTL in lettuce crop-wild hybrids: Comparing greenhouse and field experiments. *Ecology and Evolution, 4*(12), 2395–2409. https://doi.org/10.1002/ece3.1060.

Kim, I., & Lee, J. (2020). Variations in anthocyanin profiles and antioxidant activity of 12 genotypes of mulberry (*Morus* spp.) fruits and their changes during processing. *Antioxidants, 9*(3), 1–14. https://doi.org/10.3390/antiox9030242.

Kleinhenz, M. D., French, D. G., Gazula, A., & Scheerens, J. C. (2003). Variety, shading, and growth stage effects on pigment concentrations in lettuce grown under contrasting temperature regimens. *HortTechnology, 13*(4), 677–683. https://doi.org/10.21273/horttech.13.4.0677.

Koopman, W. J. M. (1999, June) Plant systematics as useful tool for plant breeders: Examples from lettuce. In: A. Lebeda & E. Krístková, E. (eds.), Eucarpia leafy vegetables '99: Proceedings of the EUCARPIA Meeting on Leafy Vegetables Genetics and Breeding (pp. 95–105). Palacký Univ., Olomouc, Czech Republic.

Koopman, W. J. M., Guetta, E., Van de Wiel, C. C. M., Vosman, B., & Van den Berg, R. G. (1998). Phylogenetic relationships among *Lactuca* (Asteraceae) species and related genera based on ITS-1 DNA sequences. *American Journal of Botany, 85*(11), 1517–1530.

Lebeda, A., Doležalová, I., & Astley, D. (2004). Representation of wild *Lactuca* spp. (Asteraceae, Lactuceae) in world genebank collections. *Genetic Resources and Crop Evolution, 51*(2), 167–174. https://doi.org/10.1023/B:GRES.0000020860.66075.f7.

Lebeda, A., Křístková, E., Kitner, M., Mieslerová, B., Jemelková, M., & Pink, D. A. C. (2014). Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. *European Journal of Plant Pathology, 138*(3), 597–640. https://doi.org/10.1007/s10658-013-0254-z.

Lee, S. K., & Kader, A. A. (2000). Preharvest and postharvest factors influencing vitamin C content of horticultural crops. *Postharvest Biology and Technology, 20*(3), 207–220. https://doi.org/10.1016/S0925-5214(00)00133-2.

Llorach, R., Martínez-Sánchez, A., Tomás-Barberán, F. A., Gil, M. I., & Ferreres, F. (2008). Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chemistry, 108*(3), 1028–1038. https://doi.org/10.1016/j.foodchem.2007.11.032.

Mallor, C., & Díaz, A. (2016). Melon germplasm characteristics, diversity, preservation and uses. In M. Walton (Ed.), Germplasm: Characteristics, Diversity and Preservation (pp. 1–26). Nova Science Publishers.

Medina-Lozano, I., Bertolín, J. R., Zufiaurre, R., & Díaz, A. (2020). Improved UPLC-UV method for the quantification of vitamin C in lettuce varieties (*Lactuca sativa* L.) and crop wild relatives (*Lactuca* spp.). *Journal of Visualized Experiments, 160*(e61440), 1–16. https://doi.org/10.3791/61440.

Mulabagal, V., Ngouajio, M., Nair, A., Zhang, Y., Gottumukkala, A. L., & Nair, M. G. (2010). *In vitro* evaluation of red and green lettuce (*Lactuca sativa*) for functional food properties. *Food Chemistry, 118*(2), 300–306. https://doi.org/10.1016/j.foodchem.2009.04.119.

Ordidge, M., García-Macías, P., Battey, N. H., Gordon, M. H., Hadley, P., John, P., … Wagstaffe, A. (2010). Phenolic contents of lettuce, strawberry, raspberry, and blueberry crops cultivated under plastic films varying in ultraviolet transparency. *Food Chemistry, 119*(3), 1224–1227. https://doi.org/10.1016/j.foodchem.2009.08.039.

Peniche-Pavía, H. A., & Tiessen, A. (2020). Anthocyanin profiling of maize grains using DIESI-MSQD reveals that cyanidin-based derivatives predominate in purple corn, whereas pelargonidin-based molecules occur in red-pink varieties from Mexico. *Journal of Agricultural and Food Chemistry, 68*(21), 5980–5994. https://doi.org/10.1021/acs.jafc.9b06336.

Petropoulos, S. A., Barros, L., & Ferreira, I. C. F. R. (2019). Editorial: Rediscovering local landraces: Shaping horticulture for the future. *Frontiers in Plant Science, 10*(126), 1–2. https://doi.org/10.3389/fpls.2019.00126.

PGR (Plant Genetic Resources) Lettuce. The lettuce gene pool. (2018). http://www.pgrportal.nl/en/Lettuce-genetic-resources-Portal.htm/ Accessed 25 July 2020.

UPOV (International Union for the Protection of New Varieties of Plants) (2019) Guidelines for the Conduct of Tests for Distinctness, Homogeneity, and Stability. Document UPOV TG/13/11 Rev, Geneva, Switzerland.

USDA (U.S. Department of Agriculture) (2016) Composition of Foods Raw, Processed, Prepared. USDA National Nutrient Database for Standard Reference, Release 28. https://data.nal.usda.gov/dataset/composition-foods-raw-processed-prepared-usda-national-nutrient-database-standard-referenc-9.

van Treuren, R., van Eekelen, H. D. L. M., Wehrens, R., & de Vos, R. C. H. (2018). Metabolite variation in the lettuce gene pool: Towards healthier crop varieties and food. *Metabolomics, 14*(11), 1–14. https://doi.org/10.1007/s11306-018-1443-8.

Viacava, G. E., Roura, S. I., Berrueta, L. A., Iriondo, C., Gallo, B., & Alonso-Salces, R. M. (2017). Characterization of phenolic compounds in green and red oak-leaf lettuce cultivars by UHPLC-DAD-ESI-QToF/MS using MSE scan mode. *Journal of Mass Spectrometry, 52*(12), 873–902. https://doi.org/10.1002/jms.4021.

Wu, X., & Prior, R. L. (2005). Identification and characterization of anthocyanins by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry in common foods in the United States: Vegetables, nuts, and grains. *Journal of Agricultural and Food Chemistry, 53*(8), 3101–3113. https://doi.org/10.1021/jf0478861.

Yousuf, B., Gul, K., Wani, A. A., & Singh, P. (2016). Health benefits of anthocyanins and their encapsulation for potential use in food systems: A review. *Critical Reviews in Food Science and Nutrition, 56*(13), 2223–2230. https://doi.org/10.1080/10408398.2013.805316.

Zhang, Y., Butelli, E., & Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Current Opinion in Plant Biology, 19*, 81–90. https://doi.org/10.1016/j.pbi.2014.05.011.

Złotek, U., Świeca, M., & Jakubczyk, A. (2014). Effect of abiotic elicitation on main health-promoting compounds, antioxidant activity and commercial quality of butter lettuce (*Lactuca sativa* L.). *Food Chemistry, 148*, 253–260. https://doi.org/10.1016/j.foodchem.2013.10.031.

Zohary, D. (1991). The wild genetic resources of cultivated lettuce (*L. sativa* L.). *Euphytica, 53*, 31–35.

CHAPTER 1. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://doi.org/10.1016/j.foodchem.2021.129864.

Figure S1. Images of the plant material used in the present study, commercial (green and red) and traditional (green and semi-red) lettuce varieties, as well as wild relatives (*Lactuca* spp.).

Table S1. Gradient conditions for the identification and quantification by UPLC of the anthocyanins present in the *Lactuca* samples included in the present study.

Time	Flow (mL/min)	%A	%C	Curve
Initial	0.3	15	85	Lineal
3.30	0.3	20	80	Lineal
3.86	0.3	30	70	Lineal
5.05	0.3	40	60	Lineal
5.35	0.3	55	45	Lineal
5.64	0.3	60	40	Lineal
5.95	0.3	95	5	Lineal
7.50	0.3	95	5	Lineal
12.00	0.3	95	5	Lineal
12.30	0.3	15	85	Lineal
17.00	0.3	15	85	Lineal

CHAPTER 2

Selection of a lettuce traditional variety with high content of vitamin C and studies of genetic diversity and genome-wide association for vitamin C

Medina-Lozano, I, Bertolín, JR, Plieske, J, Ganal, M, Gnad, H, Díaz, A (2024b). Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays. *Plant Genome* 17, e20518. https://doi.org/10.1002/tpg2.20518.

CHAPTER 2

Selection of a lettuce traditional variety with high content of vitamin C and studies of genetic diversity and genome-wide association for vitamin C

Justification

In the previous chapter we found that commercial varieties were the richest accessions in anthocyanins but the poorest in vitamin C. Anthocyanin enhancement has already been achieved in lettuce breeding programmes because it was targeted as a valuable agronomic character, the red colour of the leaves, and not because of their beneficial properties for health. Nutritional value has not been a priority in lettuce breeding and the improvement of other characters has likely been detrimental to the vitamin C content in commercial varieties. Therefore, vitamin C was the selected trait in this thesis to start a breeding programme focused on obtaining biofortified lettuce varieties. 'Lechuga del Pirineo' was the selected accession in this study because the plant with the highest vitamin C content according to the results shown in the previous chapter happened to belong to this traditional variety. So, three populations were analysed, the original variety population (S0) and two more populations (S1 and S2) coming from the selfing of the richest plant in vitamin C in two consecutive years. In addition, we wanted to get a better understanding of the genetic basis behind the vitamin C content through a marker-trait association study. To do so, a diversity panel of varieties was used to increase the genetic variation, together with the S0 'Lechuga del Pirineo' population. Presumably, the highly diverse varieties would allow us to detect new and significant association signals and the large number of less heterogenous samples from the S0 population would increase the power to detect major loci. In this way, we achieved a balance between genetic diversity and allele frequencies.

Received: 8 July 2024 | Accepted: 31 August 2024

DOI: 10.1002/tpg2.20518

The Plant Genome

ORIGINAL ARTICLE

Special Section: Enhancing Food Security through Innovative Agricultural Management

Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays

Inés Medina-Lozano[1,2] | **Juan Ramón Bertolín**[2,3] | **Jörg Plieske**[4] | **Martin Ganal**[4] | **Heike Gnad**[4] | **Aurora Díaz**[1,2]

[1]Department of Plant Sciences, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain

[2]AgriFood Institute of Aragon–IA2 (CITA-University of Zaragoza), Zaragoza, Spain

[3]Department of Animal Sciences, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain

[4]SGS Institut Fresenius GmbH TraitGenetics Section, Seeland, Germany

Correspondence
Aurora Díaz, Department of Plant Sciences, Agrifood Research and Technology Centre of Aragon (CITA), Avd. Montañana 930, 50059, Zaragoza, Spain.
Email: adiazb@cita-aragon.es; adiazb@unizar.es

Assigned to Associate Editor Lubin Tan.

Funding information
Spanish Ministry of Science and Innovation (MCIN); European Union (EU) NextGenerationEU: PRTR-C17.I1, Grant/Award Number: AGROALNEXT; MCIN; State Research Agency (AEI), Grant/Award Number: PID2022-138484OR-I00; EU; Government of Aragón: "Grupo de investigación en Fruticultura - Caracterización, Adaptación y Mejora Genética", Grant/Award Number: A12_23R; Spanish Ministry of Science, Innovation and Universities (MCIU); AEI, Grant/Award Number: I.M-L. predoctoral contract

Abstract

Lettuce (*Lactuca sativa* L.) is a source of beneficial compounds though they are generally present in low quantities. We used 40K Axiom and 9K Infinium SNP (single nucleotide polymorphism) arrays to (i) explore the genetic variability in 21 varieties and (ii) carry out genome-wide association studies (GWAS) of vitamin C content in 21 varieties and a population of 205 plants from the richest variety in vitamin C ('Lechuga del Pirineo'). Structure and phylogenetic analyses showed that the group formed mainly by traditional varieties was the most diverse, whereas the red commercial varieties clustered together and very distinguishably apart from the rest. GWAS consistently detected, in a region of chromosome 2, several SNPs related to dehydroascorbic acid (a form of vitamin C) content using three different methods to assess population structure, subpopulation membership coefficients, multidimensional scaling, and principal component analysis. The latter detected the highest number of SNPs (17) and the most significantly associated, 12 of them showing a high linkage disequilibrium with the lead SNP. Among the 84 genes in the region, some have been reported to be related to vitamin C content or antioxidant status in other crops either directly, like those encoding long non-coding RNA, several F-box proteins, and a pectinesterase/pectinesterase inhibitor, or indirectly, like extensin-1-like

Abbreviations: AA, ascorbic acid; AMOVA, analysis of molecular variance; CR, call rate; CWR, crop wild relatives; DHAA, dehydroascorbic acid; FarmCPU, fixed and random model circulating probability unification; FLD, Fisher's linear discriminant; H_O, observed heterozygosity; HomFLD, homozygous Fisher's linear discriminant; GBS, genotyping-by-sequencing; tGBS, tunable genotyping-by-sequencing; GD, gene diversity; GDSL, glycine-asparagine-serine-leucine; GWAS, genome-wide association study; IBS, identity by state; LD, linkage disequilibrium; MAF, minor allele frequency; MDS, multidimensional scaling; PCA, principal component analysis; Q, subpopulation membership coefficient; QC, quality control; QQ, quantile-quantile; SA, salicylic acid; SNP, single nucleotide polymorphism; TAA, total ascorbic acid; UPLC, ultra performance liquid chromatography; UV, ultaviolet; WPP, tryptophan-proline-proline.

protein and endoglucanase 2 genes. The involvement of other genes identified within the region in vitamin C levels needs to be further studied. Understanding the genetic control of such an important quality trait in lettuce becomes very relevant from a breeding perspective.

Plain Language Summary

Domestication and breeding have impoverished many crops, like lettuce, in terms of nutritional value. We have explored the genetic and nutritional diversity in cultivated lettuces, finding that traditional varieties are richer in both aspects than those that are commercial. The variety with the highest content in vitamin C, a traditional one, was self-pollinated to create descendants with high level of vitamin C and genetic homogeneity. Using this breeding population and the set of diverse varieties mentioned above, we have found a region of the lettuce genome associated to the amount of one of the forms of vitamin C. Among the genes in the region, there are some that have become strong putative candidates to be involved in vitamin C accumulation as they play similar roles in other crops, like wheat, tomato, and potato, among others. Knowing those genes and understanding how they work is challenging, though it will boost the breeding toward lettuce varieties biofortified in vitamin C and hence more nutritious.

1 | INTRODUCTION

Global production of lettuce (*Lactuca sativa* L.), together with chicory, exceeded 27 million tonnes, with a harvested area of 1.24 million ha in 2022 (FAOSTAT, 2022). These figures reflect the high demand among consumers. This makes lettuce an ideal candidate crop to start a breeding program aimed at improving its nutritional value and health properties as it will certainly have a strong and positive effect on a large part of the global population. Although lettuce is a source of a big range of compounds, which confer benefits to human health, such as vitamins, fiber, and phenolic compounds, they are generally only present in low quantities (USDA, 2022). In fact, when compared to other leafy vegetables, especially spinach, chard, cabbage, or watercress, lettuce is the poorest in vitamin C (USDA, 2022), which is a key indicator of the quality of fruit and vegetables. The process of wild *Lactuca* spp. domestication, which led to the present cultivated lettuce, entailed some collateral effects, like a higher susceptibility to some pathogens and pests, a loss in phytonutrients and beneficial compounds, and a decrease in genetic diversity, among others, as has also happened in other crops. The last one is particularly true in lettuce as it is a predominantly autogamous species. Breeding programs have made use of crop wild relatives (CWR) to introduce disease- and pest-resistance genes in the crop, but they are underused resources with regard to the improvement of the nutritional quality (Dempewolf et al., 2017). This could result surprising in lettuce, especially in the light of studies that have compared the content of vitamin C in different *Lactuca* spp. and found that the wild lettuce relatives are generally richer than the cultivated varieties (Medina-Lozano et al., 2021, 2024; van Treuren et al., 2018). However, linkage drag of undesirable traits linked to the genes of interest (especially those related to organoleptic or sensory attributes) could have discouraged their inclusion in modern breeding programs, in which the flavor has been prioritized over the nutritional quality in the selection. In this sense, landraces and traditional varieties have been proposed as shortcuts when compared to CWR to be used in breeding programs (Medina-Lozano & Díaz, 2021) as they do not carry detrimental and/or maladaptive variants and the reproductive barriers sometimes hindering interspecific crossings between cultivated and wild forms are avoided. Even if they do not harbor as much genetic variability as the wild germplasm, they are still more diverse than the modern commercial varieties (Flint-Garcia et al., 2023). Furthermore, in the case of vitamin C content in lettuce, they have revealed themselves to be also richer than the commercial varieties tested (Medina-Lozano et al., 2020, 2021).

Vitamin C or total ascorbic acid (TAA) is composed of ascorbic acid (AA) and its oxidation product, dehydroascorbic acid (DHAA). AA is a powerful antioxidant, which contributes to a healthy state by preventing common diseases (Granger & Eck, 2018) and must be supplemented in the diet

mainly through fruits and vegetables as humans are unable to synthesize it. DHAA is also a bioactive compound, which can be converted into AA in the human body. That is why DHAA has been suggested to serve as vitamin C reservoir under some adverse conditions, such as those causing oxidative stress, in which AA get transformed into DHAA as a consequence of its antioxidant activity (Medina-Lozano et al., 2021, 2024).

The use of genomic technologies to explore biodiversity and to associate the traits of interest to the genomic regions responsible for them (genome-wide association studies [GWAS]) are among the most used tools nowadays in early stages of crop biofortification. The predominant markers to assess genomic diversity and genetic structure are SNPs (single nucleotide polymorphisms), thanks to their abundance in all species, and the availability of high-throughput SNP genotyping platforms in many food crops (Medina-Lozano & Díaz, 2022). At the moment, more than 70K SNP data are available at the Lettuce Genome Database (LettuceGDB, https://www.lettucegdb.com). Thousands of SNPs have been obtained and used to explore *Lactuca* diversity, mainly among lettuce cultivars but also including CWR, with different technologies such as (i) microarrays like Illumina GoldenGate (S. J. Kwon et al., 2012) or Affymetrix GeneChip (Stoffel et al., 2012) assays and (ii) next-generation sequencing like single primer enrichment technology (Tripodi et al., 2023), genotyping-by-sequencing (GBS) (J. S. Park et al., 2022), tunable GBS (tGBS) (S. Park et al., 2021; Simko, 2023), and even whole-genome resequencing (Wei et al., 2021), which has rendered 179 million SNPs, as well as other types of variants. Some of these studies have also inquired into lettuce domestication history using the diversity panel of 445 *Lactuca* accessions firstly characterized by Wei et al. (2021).

There are a substantial number of works on GWAS in lettuce regarding different types of traits, like those related to tolerance to biotic or abiotic stress, development, nutrient efficiency use, postharvest behavior, and morphology, among others. However, GWAS of characters related to lettuce health-promoting properties or nutritional value are not common. Interestingly, one of the few types of metabolites targeted by GWAS in lettuce, anthocyanins, are bioactive compounds with antioxidant activity (as it is the case of vitamin C), though they were originally addressed as morphological traits, either as a qualitative character, leaf color (L. Zhang et al., 2017), or measuring their content in leaves (Tripodi et al., 2023; Wei et al., 2021). Other GWAS have been carried out to dissect the genetic basis of primary metabolite content, partly responsible (together with others) for the nutritional value of lettuce (W. Zhang et al., 2020). Information on both types of compounds, those with positive effects on human health and nutrients, is actually very helpful to pursue the biofortification of the crop.

To date, different approaches have been adopted to enhance vitamin C content in lettuce, either conventional, by supplying UV (ultra-violet)-B radiation (H. Zhou et al., 2023) or applying minerals (Dyląg et al., 2023) to the plants, or based on genetic engineering techniques (Guo et al., 2013). Both present disadvantages such as temporary effects, in the case of conventional strategies, and legislative issues in some countries, in the case of modern genetic engineering methods. However, classic genetic breeding has also limitations, for instance, only the variability present in plants from sexually compatible groups can be used.

In this work, we explore and exploit the diversity, both genetic and nutritional, within the cultivated lettuce, including traditional varieties. With the richest accession in vitamin C, which happened to be a traditional variety, breeding populations were built. Those, together with a diversity panel of commercial and traditional varieties, were used to find a genomic region associated to vitamin C content and identify putative candidate genes to boost the breeding toward lettuce varieties biofortified in vitamin C and thereupon healthier and more nutritious.

Core Ideas

- Lettuce is ideal to be biofortified as it is nutritionally poor (i.e., vitamin C) but highly demanded by consumers.
- Traditional varieties harbor a great genetic diversity, essential for the breeding of this autogamous species.
- These are the first genetic associations with dehydroascorbic acid content found in lettuce: a 5.1Mb region in chromosome 2.
- High linkage disequilibrium values were only found between the lead single nucleotide polymorphism (SNP) and other significantly associated SNPs.
- Some of the candidate genes in the region of interest are involved in vitamin C metabolism in other crops.

2 | MATERIALS AND METHODS

2.1 | Plant material and trait evaluation

The present study was divided into two different parts: (i) a genetic diversity analysis and (ii) a GWAS of vitamin C content.

First, a total of 21 lettuce accessions were used in the genetic diversity analysis (Table 1). These included 10 commercial varieties (4 green and 6 red) and 11 Spanish

TABLE 1 Description of the plant material and the assays carried out in the present study.

Variety or population name	Assay (No. of SNPs)	Description	Type[a]	Origin	Source[b]	Accession number
'Begoña'	Genetic diversity (13,026) and GWAS (9,242)	Green commercial	Batavia	Spain	Ramiro Arnedo Semillas S.A.	–
'Dolomiti G12'	Genetic diversity (13,026) and GWAS (9,242)	Green commercial	Gem	Spain	Ramiro Arnedo Semillas S.A.	–
'Lechuga de Beceite'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ2006
'Lechuga de Bureta'	Genetic diversity (13,026) and GWAS (9,242)	Semi-red traditional	Cos	Spain	BGHZ	BGHZ4927
'Lechuga de Ensalada'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ2031
'Lechuga de Híjar'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ0529
'Lechuga de Subías'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ1852
'Lechuga del Pirineo'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ2229
'Lechuga del Valle de Tena'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Butterhead	Spain	BGHZ	BGHZ1850
'Lengua de Buey'	Genetic diversity (13,026) and GWAS (9,242)	Green traditional	Cos	Spain	BGHZ	BGHZ2004
'Likarix'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Frisée d'Amérique	Netherlands	CGN	CGN24522
'Lollo Rosso'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Lollo	Italy	CGN	CGN09385
'Morada de Belchite'	Genetic diversity (13,026) and GWAS (9,242)	Semi-red traditional	Cos	Spain	BGHZ	BGHZ0527
'Morada de Bernués'	Genetic diversity (13,026) and GWAS (9,242)	Semi-red traditional	Batavia	Spain	BGHZ	BGHZ2097
'Morada de Sorripas'	Genetic diversity (13,026) and GWAS (9,242)	Semi-red traditional	Cos	Spain	BGHZ	BGHZ2026
'Nestorix'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Lollo	Netherlands	CGN	CGN24712
'Red Sails'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Lollo	Germany	CGN	CGN19014
'Revolution'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Lollo	Netherlands	CGN	CGN20714
'Romana Inverna'	Genetic diversity (13,026) and GWAS (9,242)	Green commercial	Cos	Spain	BGHZ	BGHZ3604
'Romired'	Genetic diversity (13,026) and GWAS (9,242)	Red commercial	Lollo	Netherlands	CGN	CGN24713
'Winter Crop'	Genetic diversity (13,026) and GWAS (9,242)	Green commercial	Butterhead	Hungary	CGN	CGN05853
'Lechuga del Pirineo' S0	GWAS (9,242)	205 plants (green traditional)	Cos	Spain	–	–
'Lechuga del Pirineo' S1	–[c] (9,242)	239 full-sibs (green traditional)	Cos	Spain	–	–
'Lechuga del Pirineo' S2	–[c] (9,242)	179 full-sibs (green traditional)	Cos	Spain	–	–

Abbreviations: GWAS, genome-wide association study; SNP, single nucleotide polymorphism.
[a]According to the International Union for the Protection of New Varieties of Plants (UPOV, 2021).
[b]BGHZ: Vegetable Germplasm Bank of Zaragoza (Spain); CGN: Centre for Genetic Resources (Wageningen, Netherlands).
[c]Populations originally built but not included in the GWAS because resulted monomorphic for all the markers.

traditional varieties (7 green and 4 semi-red), representing six types of lettuce attending to their morphology: Batavia, Butterhead, Cos, Frisée d'Amérique, Gem, and Lollo (UPOV, 2021).

Second, for the GWAS, the plant material originally included target populations coming from a breeding program aimed at enhancing the vitamin C content in lettuce. The plant with the highest vitamin C content according to a previous study with *Lactuca* germplasm, these 21 lettuce varieties (among others) and some wild relatives, belonged to the traditional variety 'Lechuga del Pirineo' (Medina-Lozano et al., 2021). From this one, three populations were analyzed, the original variety population (S0) made up of 205 plants (presumably coming from seeds of different plants supplied by a germplasm bank, BGHZ, Table 1), and two self-pollination generations, S1 and S2, composed of 239 and 179 plants, respectively, obtained by selfing the plant with the highest vitamin C content in each generation to increase both the metabolite concentration and the genetic homogeneity (Table 1). As lettuce is a predominantly autogamous species, the genetic variability was already negligible in S1 and, obviously, in S2 (Table S1), so they were not used in association analysis. To increase GWAS's power, the needed heterozygosity was incorporated by using the 21 varieties mentioned above from the diversity panel (Table 1), as recommended by Hamazaki et al. (2020).

For both analyses, genetic diversity (21 varieties) and GWAS (21 varieties and 205 individuals of 'Lechuga del Pirineo' S0), plants were grown in pots (30 × 25 cm and 11.7 L volume) with a mix of black and blonde peat (1:1) in a greenhouse at Agrifood Research and Technology Centre of Aragon (CITA, Zaragoza, Spain). Plants for the diversity study were cultivated in winter 2018/2019, and S0, S1, and S2 'Lechuga del Pirineo' populations were cultivated in winters 2020/2021, 2021/2022, and 2022/2023, respectively. After a period ranging from 2.5 to 4 months, depending on the accession and the population, leaves were harvested as described in the next section and immediately frozen with liquid nitrogen and kept at −80°C.

For the GWAS, the two forms of vitamin C, AA and DHAA, as well as the total content, TAA, were quantified in samples consisting of both inner and outer leaves from 644 plants (205 S0, 239 S1, 179 S2, and 21 varieties) by UPLC (ultra performance liquid chromatography) according to the method described by Medina-Lozano et al. (2020). Briefly, the extraction was conducted using 50 mg of finely powdered lyophilized samples with 5 mL of a solution of 8% acetic acid (v/v), 1% meta-phosphoric acid (w/v), and 1 mM EDTA (ethylenediaminetetraacetic acid). The mixture was vortexed for 5 s, shaken for 10 min at 2000 rpm, and then sonicated for 20 min at room temperature, and centrifuged at 4000 × *g* for 10 min at 4°C. The supernatant was filtered through a 0.22-µm regenerated cellulose filter (Agilent). The filtrate (Extract 1, E1) contained both AA and DHAA. Two 200-µL aliquots of E1 were used to determine (i) AA directly and (ii) TAA by reducing DHAA to AA adding 200 µL of a reduction solution (40 mM DTT [dithiothreitol] with 0.5 M Tris pH 9.0) and stopping the reaction after 30 min with 200 µL of 0.4 M sulfuric acid to obtain Extract 2 (E2). This last step is needed because DHAA absorptivity in the UV range of the spectrum is too low to be measured directly. A volume of 5 µL of E1 and E2 diluted with ultrapure water (1:4 v:v) was injected in a liquid chromatographer UPLC H-Class with an HSS T3 column (150 mm × 2.1 mm × 1.8 µm). The total running time was 3 min and the temperature of the samples and the column was programed at 5°C and 30°C, respectively. The wavelength of the Acquity UPLC Photodiode Array eλ detector was set at 245 nm. The mobile phase consisted of 2% methanol and 98% ultrapure water pH 2.0 acidified with formic acid at a flow rate of 0.3 mL min^{-1} in isocratic mode. For quantification of AA and TAA contents, a calibration curve from 0.5 to 25.0 µg mL^{-1} of the commercial L-ascorbic acid (≥99.9% purity, Sigma-Aldrich) was built. DHAA content was calculated by subtracting AA from TAA.

Effects of the generation (S0, S1, and S2) on vitamin C content (AA, DHAA, and TAA) were tested with an analysis of the variance by Kruskal–Wallis test and *post hoc* Dunn's test for mean comparison using a Bonferroni corrected $\alpha = 0.017$.

2.2 | DNA extraction, SNP genotyping, and quality control

DNA was extracted from young leaves of 644 plants (205 S0, 239 S1, 179 S2, and 21 varieties) as described in Doyle and Doyle (1990) with the following modifications (Díaz et al., 2017): 0.2% β-mercaptoethanol was added together with the 2% CTAB (hexadecyltrimethylammonium bromide) buffer, the washing buffer consisted of 76% ethanol and 10 mM ammonium acetate, and 0.2 µL of 10 mg × mL^{-1} RNAse A (Invitrogen) was added to 30 µL of milliQ water to dissolve the pellet. DNA quality control (QC) was carried out via electrophoresis in 1% agarose gels. DNA concentration was measured fluorometrically using a Quantiflour-ST (Promega GmbH).

DNA sample genotyping was performed using the lettuce 40K Axiom and 9K Infinium arrays developed by the SGS Institut Fresenius GmbH TraitGenetics Section, containing 41,975 and 9,381 SNPs, respectively. Axiom array was scanned with a GeneTitan Scan Instrument (Thermofisher Scientific) followed by data analysis with Axiom Analysis Suite software v5.1.1.1 (Thermofisher Scientific) using the default settings. SNPs were filtered by values of call rate (CR) >90%, Fisher's linear discriminant (FLD) >5,

and homozygous FLD (HomFLD) >10. Infinium array was scanned with an iScan system (Illumina) followed by data analysis with GenomeStudio 2.0 (Illumina).

Several QC steps were undertaken with the whole set of markers to be used in both studies, diversity and GWAS. After compiling all data with GenomeStudio software, sample clustering was manually optimized, and SNP markers were filtered by values CR >90%. To improve the data quality, a comparison between common markers from both arrays that met the QC criteria just described was carried out. The markers rendering incongruous genotypes were discarded. Another quality filter was applied by discarding the markers with spurious genotypes according to the pedigree in the 'Lechuga del Pirineo' populations. Then, only polymorphic SNPs in the corresponding set of samples for each analysis were selected: 13,026 markers in the genetic diversity study and 9,242 in the GWAS (Tables S1 and S2).

2.3 | Genetic and genomic analyses

2.3.1 | Population structure and genetic diversity and relationships

The population structure of the varieties in the diversity panel was analyzed carrying out a simulation using the Bayesian algorithm and the admixture model in STRUCTURE v2.3.4 (Pritchard et al., 2000). A burn-in period of 100,000 cycles followed by 100,000 Markov chain Monte Carlo iterations was tested with a number of subpopulations (K) set from one to six. Ten independent runs per K value were performed. The optimal number of K was inferred applying the ΔK method (Evanno et al., 2005). The analysis was repeated with the same parameters for K = 2 and K = 3. Data were plotted using the web application Structure Plot v2.0 (Ramasamy et al., 2014).

Alternatively, both a principal component analysis (PCA) and a multidimensional scaling (MDS), also known as principal coordinate analysis, were performed to visualize patterns of diversity using TASSEL v5.2 (Bradbury et al., 2007) and the software JMP v17.2 for Windows (SAS Institute Inc.) was used to plot the results.

To explore the genetic relationships between the 21 varieties included in the diversity study, a matrix of genetic distances was created from the genotypic data using the IBS (identity by state) method in TASSEL v5.2 software. A phylogenetic network was built based on the genetic distance matrix with the NeighborNet method using the SplitsTree App v6.3.12 software (Huson & Bryant, 2006).

The genetic differentiation was assessed based on the number of distinct populations obtained in the genetic structure analysis (K = 3). Wright's fixation index F_{ST} among populations was calculated. An analysis of molecular variance (AMOVA) was conducted to detect genetic variation among and within populations, as well as within individuals. Summary statistics including gene diversity (GD) and observed heterozygosity (H_O) were calculated for each subpopulation. All the above statistical analyses were performed using PowerMarker 3.25 software (K. Liu & Muse, 2005).

2.3.2 | Linkage disequilibrium, GWAS, and identification of candidate genes

In the GWAS subset of markers, SNPs were filtered for minor allele frequency (MAF) >0.01. This value was not arbitrarily chosen but justified by the nature of the samples. Since most plants belonged to the same variety, 'Lechuga del Pirineo' (205 S0 plants out of 226, Table 1), minor alleles present in that accession were overrepresented and vice versa, the frequency of non-rare alleles in the 21 varieties diversity panel was diluted in the whole set of samples. With the conventional threshold MAF >0.05, a variant should have to be present in more than 11 samples for not being filtered out, which means to be in more than half of the samples of the diversity panel, which obviously is not a minor allele. MAF >0.01 ensures that only variants appearing less than twice were eliminated. Chromosomal and physical SNP positions were determined based on lettuce reference genome Lsat_Salinas_v11 (GCF_002870075.4).

Linkage disequilibrium (LD) between SNPs on each chromosome was calculated through pairwise correlation coefficients (r^2) with an LD window size of 50 sites around each marker. LD decay was determined by plotting r^2 values against the physical distance of the SNPs and then plotted in Rstudio (Rstudio Team, 2020).

Multi-locus GWAS was conducted using GAPIT3 (Wang & Zhang, 2021) in Rstudio. Fixed and random model Circulating Probability Unification (FarmCPU) was used considering both kinship and structure of the samples (X. Liu et al., 2016). Kinship was assessed with GAPIT3 package. Three different methods of measuring the structure were used: a PCA performed with TASSEL v5.2 software, an MDS analysis also conducted in TASSEL v5.2 from the genetic distance matrix calculated using the IBS method, and subpopulation membership coefficients (Q) obtained with STRUCTURE software using the linkage model at K = 3 and the rest of the settings as described earlier. The threshold p-value and $-\log_{10}(p\text{-value})$ were determined using the Bonferroni correction with genome-wide significance level of $\alpha = 0.05$ and taking into account the total number of SNPs (9,242) as follows: $p = 0.05/9{,}242 = 5.41 \times 10^{-6}$ and $-\log_{10}(5.41 \times 10^{-6}) = 5.27$. Manhattan and quantile–quantile (QQ) plots for GWAS results were obtained using CMplot package (Yin et al., 2021) in Rstudio.

A local LD block analysis was conducted in the region of associated SNPs using the LDBlockShow software (Dong

et al., 2021) in Linux environment. The most significantly associated SNP in the GWAS analysis was set as the lead SNP.

Finally, putative candidate genes for vitamin C content (AA, DHAA, and TAA) in the region were identified in the annotated version of *L. sativa* genome (Lsat_Salinas_v11, GCF_002870075.4) available at the NCBI database.

3 | RESULTS AND DISCUSSION

3.1 | Assessment of population structure and genetic diversity and relationships

A total of 13,026 polymorphic SNPs were used to analyze the population structure of a panel of 21 lettuce varieties as well as the genetic diversity and relationships.

The best K estimated by calculating ΔK using the method described by Evanno et al. (2005) indicated that the optimal number of subpopulations for the 21 lettuce varieties was three, corresponding to the highest peak in the ΔK plot (Figure 1A; Figure S1). The first population (Pop. 1) grouped five varieties, consisting of red commercial lettuces exclusively, that came from the Netherlands and Italy. Most red commercial lettuces were entirely assigned to Pop. 1, except for 'Revolution' that still showed a Q1 of 0.69. The second population (Pop. 2) was the largest and the most diverse group, including 12 lettuce varieties. It comprised both commercial and traditional varieties, which were mainly green and semi-red (there was only one red commercial variety), most of them coming from Spain. Finally, the third population (Pop. 3) consisted of four green traditional varieties native from a small geographic region (Figure S2). Among them, three out of four were 100% assigned to Pop. 3. The other one ('Lengua de Buey') showed a high level of admixture with a high Q2 value (0.41). Except for two commercial varieties, one red and one green, both from European countries other than Spain, Pop. 2 members were closer to Pop. 3 than to Pop. 1, especially some of the traditional varieties. This, together with the fact that Pop. 3 was composed exclusively of traditional varieties, suggests that Pop. 3 could really be a subgroup of Pop. 2. Indeed, when two populations were assumed (K = 2), all the varieties of Pop. 3 became part of Pop. 2 (close genetic relatedness between these two subgroups), whereas Pop. 1 remained unchanged (Figure 1A).

To evaluate genetic relationships, a phylogenetic network was obtained from genetic distances calculated using the IBS method (Figure 1B). The most clearly differentiated cluster comprised the five red commercial varieties from Pop. 1 in the structure analysis (Figure 1A,B). The four green traditional varieties of Pop. 3 also clustered together in the phylogenetic network, again with three varieties closely related and 'Lengua de Buey' a bit apart from them (Figure 1B). This cluster was not as distant from Pop. 2 as the one formed by the varieties of Pop. 1, as previously observed in the genetic structure studies. Within varieties of Pop. 2, some clustered together in a similar way as they did in the structure analysis like 'Red Sails' and 'Begoña' or 'Winter Crop', 'Lechuga del Valle de Tena' and 'Dolomiti G12' (Figure 1A,B). Overall, the phylogenetic network was in agreement with the genetic population structure. The results from STRUCTURE were also validated with a PCA and an MDS analysis (Figure 1C,D). The first two components were represented in the PCA, explaining 20.07% and 12.36% of the variation, respectively. Likewise, the first two dimensions were represented in the MDS analysis, capturing 40.74% and 23.31% of variation, respectively. In both analyses, the 21 varieties were divided in three groups, as observed before. The red commercial varieties of the Pop. 1 were clearly grouped separately from the rest by the PC1 and Dim1 from the PCA and MDS, respectively (Figure 1C,D). The green traditional varieties of Pop. 3 were completely separated in the PCA (Figure 1C), whereas in the MDS, they clustered together but overlapped completely with the 95% confidence ellipsis of Pop. 2 (Figure 1D), reflecting the proximity between these two populations.

The results from the four approaches to assess the population structure were highly consistent. Taking all of them together, we could conclude that the 21 lettuce varieties consisted of three main groups. On the one hand, Pop. 1 was the most distinguishable population in all cases, as expected, since all the varieties belonging to similar types have the same leaf color and come from a common geographical region in Europe (Table 1). On the other hand, Pop. 2 was the most diverse group and results suggested that Pop. 3 could be a subgroup of Pop. 2, as mentioned above. This reinforces the idea of traditional varieties harboring a great diversity (Medina-Lozano & Díaz, 2021), being separated in different subpopulations, even when they all come from a small area, as it is the case of all the traditional varieties in Pop. 2 and the whole Pop. 3 (Figure S2). In previous studies carried out to assess the genetic variability of *Lactuca* spp., diversity panels were mainly composed of commercial varieties, advanced breeding lines and breeding populations, like recombinant inbred lines, and lettuce wild relatives (J. S. Park et al., 2022; Peng et al., 2022). Our results suggest that traditional varieties might be an important source of genetic variability in future breeding programs as selection has had a detrimental effect on the crop diversity (J. S. Park et al., 2022). Other studies on genetic structure and phylogenetic relationships conducted in different lettuce accessions found that varieties could group together according to their morphological type (S. Kwon et al., 2013; S. Park et al., 2021; Tripodi et al., 2023). We observed something similar for Pop. 1, composed mainly of Lollo lettuces, as well as for Pop. 3, consisting exclusively of Cos varieties (Table 1; Figure 1). Pop. 2 also included

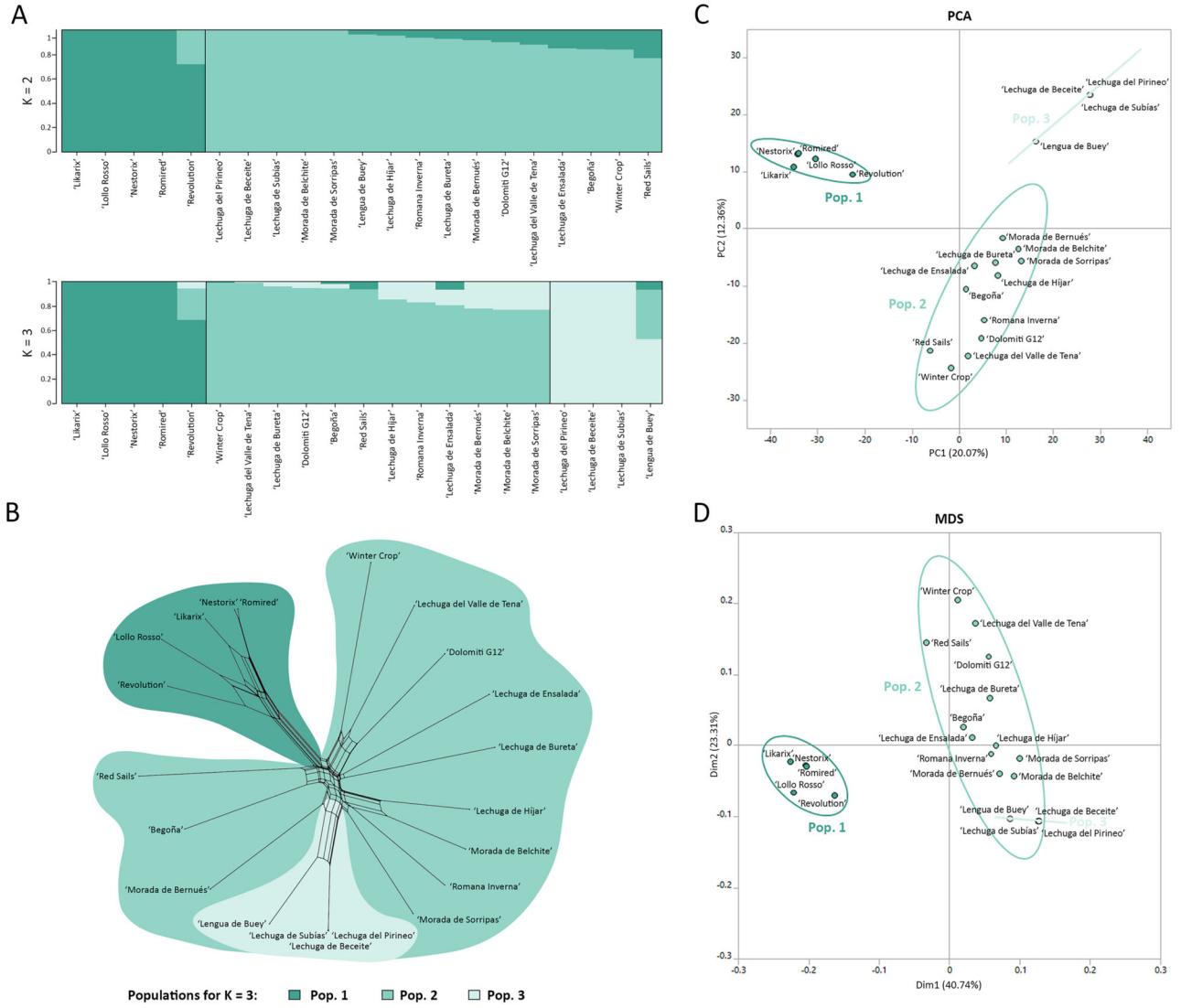


FIGURE 1 Genetic structure and relationships of 21 lettuce varieties genotyped with 13,026 single nucleotide polymorphisms (SNPs). (A) Sample membership to each of the clusters obtained with the STRUCTURE software (K = 2, 3). (B) Network built using the NeighborNet method. (C) Representation of the first two axes from a principal component analysis (PCA) and (D) a multidimensional scaling (MDS) analysis. Confidence ellipses at 95% are drawn.

Cos lettuces, what again supports the idea of Pop. 3 being a subgroup of Pop. 2. Nevertheless, more factors seemed to be influencing the population structure in our study, such as the geographical origin, especially in the case of the traditional varieties (Figure S2).

To assess the genetic variation among the populations obtained from the structure analysis, pairwise F_{ST} values were calculated, with higher F_{ST} values indicating stronger differentiation (Weir & Hill, 2002). The lowest F_{ST} values were observed between Pop. 1 and Pop. 2, followed by Pop. 2 and Pop. 3 (0.36 and 0.43, respectively) (Table 2). This could be explained by the fact that the sort of accessions in common in Pop. 1 and Pop. 2 are less diverse (commercial varieties), whereas the shared type of accessions between Pop. 2 and Pop. 3 are those with greater diversity (traditional varieties).

TABLE 2 Pairwise F_{ST} values for population differentiation.

Population	F_{ST} (Pop. 1)	F_{ST} (Pop. 2)	F_{ST} (Pop. 3)
F_{ST} (Pop. 1)	0.00		
F_{ST} (Pop. 2)	0.36	0.00	
F_{ST} (Pop. 3)	0.63	0.43	0.00

So, Pop. 2 occupied an intermediate position which agrees with being the group with more admixture (Figure 1), including varieties of all colors, types, and origins. The highest differences were obtained between Pop. 1 (exclusively formed by red commercial varieties) and Pop. 3 (only composed by green traditional varieties) ($F_{ST} = 0.63$) (Table 2), as expected since both groups were quite homogeneous and did not share

TABLE 3 Genetic variation at population (Pop.) and individual levels.

Source of variation		Sum of squares	Variation (%)
Among populations		50,660.71	28.97
Within populations	Pop. 1	17,142.40	9.80
	Pop. 2	93,025.33	53.20
	Pop. 3	6,469.75	3.70
Within individuals	Pop. 1	2,176.00	1.24
	Pop. 2	5,168.00	2.96
	Pop. 3	233.00	0.13
Total		174,875.19	100.00

Note: Analysis of molecular variance (AMOVA) using the genotyping data coming from 21 lettuce varieties.

TABLE 4 Genetic variability of 21 lettuce varieties grouped in the populations (Pops.) obtained by the structure analysis and measured as mean genetic diversity (GD) and observed heterozygosity (H_O).

Population	GD	H_O
Pop. 1	0.148	0.033
Pop. 2	0.314	0.033
Pop. 3	0.064	0.004

any common characteristics between them, such as leaf color, type, or geographical origin.

The genetic differentiation at both population and individual levels was also analyzed conducting an AMOVA. The results indicated that 28.97% of the genetic variation was observed among populations, 66.70% within populations, and 4.33% within individuals (Table 3). Of the three populations, the highest percentage of variation was found within Pop. 2 (53.20%), followed far behind by Pop. 1 (9.80%), and then by Pop. 3 (3.70%) (Table 3). Within individuals, the same pattern as within populations was obtained, the highest variation was observed within individuals of Pop. 2, followed by those of Pop. 1 and Pop. 3 with values of 2.96%, 1.24%, and 0.13%, respectively (Table 3). These results confirmed that Pop. 2 was the most diverse group and that varieties within Pop. 3 exhibited a strong similarity among them, consistently with the structure and phylogenetic analyses. In the same way, GD was higher in Pop. 2 (0.314), followed by Pop. 1 (0.148), and then by Pop. 3 (0.064) (Table 4). Finally, H_O was 0.033 in both Pop. 1 and Pop. 2, and 0.004 in Pop. 3 (Table 4). These low H_O values shown by the three populations could be explained by the fact that lettuce is a predominantly autogamous crop, so homozygous genotypes are very common. Tripodi et al. (2023) also obtained average H_O values below 4% for the groups of lettuce varieties included in their study. Despite those GD and H_O values, we observed substantial genetic variability within the set of 21 lettuce varieties, especially in the subset of traditional varieties, which is of great importance in breeding programs aimed at developing new varieties with desired traits.

3.2 | LD and association analyses, and identification of putative candidate genes

LD, defined as the nonrandom association of alleles at different loci in a given population, plays an important role in association studies. Therefore, LD was estimated between pairs of SNPs along all chromosomes. The LD decay to half r^2 ranged from 8.84 to 21.79 Mb, with a global average value of 14.92 Mb (Figure S3). Although differences among the chromosomes were observed, LD decay was overall slow, as expected in a predominantly autogamous species like *L. sativa*. Similarly, a long average value of LD decay (9.6 Mb) has been previously reported in lettuce (Simko et al., 2022). In general, autogamous species have lower recombinant rates than those that are allogamous and show slower LD decay. Genetic breeding could be favored from this LD extension, as trait-marker associations could be more probably identified than in the case of species with faster LD decay, in which the regions in LD are shorter and hence could contain a lower number of markers (Flint-Garcia et al., 2003).

In the current study, a total of 9,242 markers were used for carrying out a GWAS of vitamin C content in lettuce. 'Lechuga del Pirineo' was selected to obtain breeding populations since a plant belonging to this traditional variety was the richest in vitamin C in a previous study carried out within our group (Medina-Lozano et al., 2021). Vitamin C content (AA and DHAA) was measured in three 'Lechuga del Pirineo' populations, the original S0, and two generations coming from self-fecundation of the plant with the highest content in vitamin C, S1 and S2 (Table S3; Figure S4). The differences were highly significant ($p < 0.001$) when the average contents of AA (H: 154.38), DHAA (H: 364.97), and TAA (H: 120.73) were compared among generations. Interestingly, even though TAA was lower in S2, the content of the most active form of vitamin C (AA) increased with the two rounds of

selfing (Figure S4). Genetic homogeneity was already reached at S1 for our set of markers, so polymorphisms were only detected in S0 (Table S1). Due to the limited genetic variability of the target population (S0), a diversity panel consisting of 21 lettuce varieties was included in the GWAS, as recommended by Hamazaki et al. (2020). In this way, the whole set of samples consisted of subpopulations with different genetic backgrounds (Figure 1A), which is desirable to detect new variants, and provides the necessary genetic variability and a better genome coverage with polymorphic markers. The variability in terms of vitamin C content was also a bit higher in the diversity panel (152–424 mg × 100 g^{-1}) when compared to 'Lechuga del Pirineo' S0 population (184–424 mg × 100 g^{-1}) (Table S3).

To search for marker associations with the vitamin C content in lettuce, the GWAS was performed on DHAA, AA, and TAA contents using the FarmCPU model. FarmCPU is a multi-locus method able to control false positives incorporating both population structure and kinship, preventing overfitting by the estimation of the associations through fixed and random effect models (X. Liu et al., 2016). Controlling population structure effects is essential in GWAS (Tibbs Cortes et al., 2021), as such, different approaches have been undertaken in the present study. Three different methods were used to assess population structure: obtaining Q from the analysis with STRUCTURE software, MDS, and PCA. Several significantly associated SNPs to DHAA content were found in the same region of chromosome 2 using the three methods to elucidate the aforementioned population structure (Figure 2A), though no significant associations were obtained with either AA or TAA (Figures S5 and S6). Based on the conservative Bonferroni correction, genome-wide threshold was set at 5.27 [$-\log_{10}(5.41 \times 10^{-6})$], as explained earlier. The highest number of significantly associated SNPs was obtained using the PCA method, with a total of 17 SNPs (Figure 2A; Table 5). Using Q values and MDS, five and four associated SNPs were identified, respectively. Not only the number of associated SNPs but also the significance levels were higher using PCA, that ranged from 4.30×10^{-6} to 7.48×10^{-12}, in comparison with the Q values and MDS (2.42×10^{-6} to 1.81×10^{-7} and 4.33×10^{-6} to 7.92×10^{-7}, respectively). Despite the differences, results from the three approaches were consistent given that the significant SNPs found when using Q values and MDS were among the most significant SNPs obtained using PCA, including the lead SNP, which is the most significant one (Table 5). In a previous study carried out on baby leaf lettuce for postharvest and developmental traits, a higher number of significant associations were also reported when the PCA structure was used compared to Q coefficients (Sthapit Kandel et al., 2022).

All the significant SNPs were not only located on chromosome 2, but in a particular region of the chromosome, as mentioned earlier. Therefore, LD was analyzed in detail in

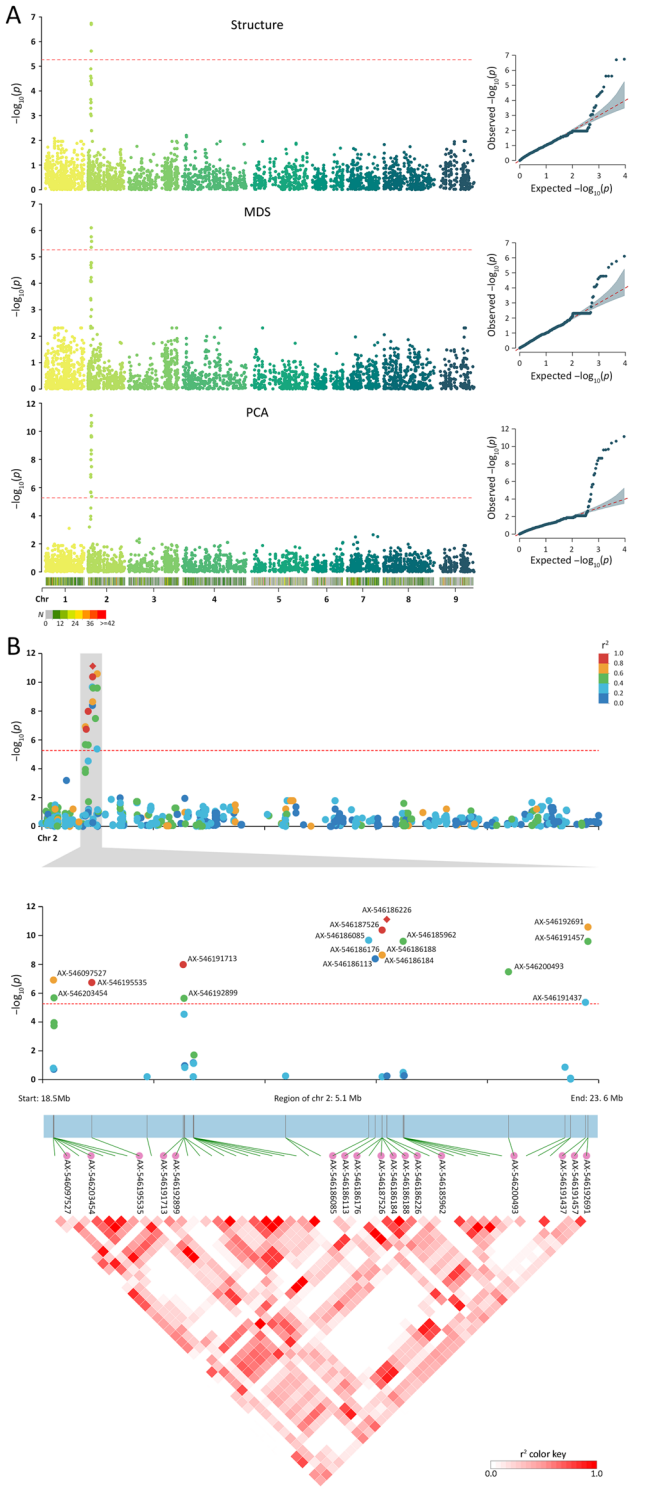


FIGURE 2 Genome-wide association study (GWAS) and linkage disequilibrium (LD) for dehydroascorbic acid (DHAA) content in lettuce samples (21 varieties and 205 plants of 'Lechuga del Pirineo' S0) genotyped with 9,242 single nucleotide polymorphisms (SNPs). (A) Manhattan and quantile–quantile (QQ) plots using the Fixed and random model Circulating Probability Unification (FarmCPU) considering both kinship and the structure of the samples measured with three methods: subpopulation membership coefficients (Q) obtained with STRUCTURE software using the linkage model and

(Continues)

TABLE 5 SNP (single nucleotide polymorphism) markers significantly associated with DHAA (dehydroascorbic acid) content in 226 lettuce plants using three approaches to assess the population structure: PCA (principal component analysis), MDS (multidimensional scaling) and STRUCTURE software.

			p-value		
SNP	**Chromosome**	**Position (bp)**	**PCA**	**MDS**	**Structure**
AX-546097527	2	18,549,215	1.21×10^{-7}	–	–
AX-546203454	2	18,553,852	2.09×10^{-6}	–	–
AX-546195535	2	18,908,075	1.81×10^{-7}	–	–
AX-546191713	2	19,764,398	1.03×10^{-8}	–	–
AX-546192899	2	19,772,368	2.26×10^{-6}	–	–
AX-546186085	2	21,502,354	2.10×10^{-10}	–	–
AX-546186113	2	21,564,067	4.08×10^{-9}	–	–
AX-546186176	2	21,628,199	2.23×10^{-9}	–	2.42×10^{-6}
AX-546187526	2	21,628,541	4.12×10^{-11}	1.73×10^{-6}	1.81×10^{-7}
AX-546186184	2	21,629,055	2.23×10^{-9}	–	2.42×10^{-6}
AX-546186188	2	21,629,299	2.23×10^{-9}	–	2.42×10^{-6}
AX-546186226	2	21,672,408	7.48×10^{-12}	7.92×10^{-7}	2.00×10^{-7}
AX-546185962	2	21,825,828	2.48×10^{-10}	4.33×10^{-6}	–
AX-546200493	2	22,812,807	3.27×10^{-8}	–	–
AX-546191437	2	23,533,970	4.30×10^{-6}	–	–
AX-546191457	2	23,557,223	2.52×10^{-10}	–	–
AX-546192691	2	23,557,727	2.55×10^{-11}	2.58×10^{-6}	–

FIGURE 2 (Continued)
K = 3, MDS (multidimensional scaling) analysis, and PCA (principal component analysis). Statistical significance threshold is shown with the horizontal line ($-\log_{10}(0.05/9242) = 5.27$). Chr: chromosome; *N*: marker density. (B) Zoom of the region harboring the significantly associated SNPs to DHAA and the squared correlation coefficients (r^2) of each marker with the lead SNP. LD patterns for the 17 SNPs significantly associated to DHAA are shown. Triangle plot depicts the LD structure of the associated region.

the region harboring the associated SNPs that covered from 18.5 to 23.6 Mb. Among them, 12 markers showed a high LD with the lead SNP ($r^2 > 0.5$) (Figure 2B). Interestingly, high LD values of the lead SNP with the rest of the SNPs present in that region were exclusively found among the significant ones. Nevertheless, a higher significance in the set of associated SNPs did not necessarily imply a higher LD with the lead SNP. To illustrate this, the second most significant SNP, and the ones following it, did not show a sequentially decreasing LD with the lead SNP (Figure 2B). These differences in the LD values observed for the most significant SNPs among the associated ones might mean that there is more than one polymorphism responsible for (or linked to) the mutation that influences DHAA content. This makes sense because vitamin C content is a complex trait controlled by multiple genes. Similarly, the associated SNPs that were physically closest to the lead SNP were not necessarily the ones with the highest LD values (Figure 2B). One reason to explain this might be that the breeding for a phenotype of a particular trait, which is controlled by more than one locus, may have resulted in the selection of variants in those loci, which will then be in high LD although they can be physically distant (Flint-Garcia et al., 2003). In addition, the significantly associated SNPs might not be within the gene responsible for the phenotypic variation observed but be in high LD with it. For this reason, genes that are in this region must be explored to find candidates related to changes in lettuce DHAA content, as not all genes were covered with SNPs.

A total of 84 genes were found in the region comprised between 18.5 and 23.6 Mb of chromosome 2 (Figure 3; Table 6). The 17 significant SNPs were within the sequence of 12 of those genes. In particular, the lead SNP was located in the uncharacterized gene of a long non-coding RNA (lncRNA) (LOC111920743), which is a class of RNA molecules of over 200 nucleotides length with none or limited coding capacity. They have been intensively studied in recent years and are known to affect gene expression in many biological processes in plants, as reviewed by J. Liu et al. (2015). Therefore, the lncRNA containing the lead SNP in the current study could be regulating the expression of genes that participate in DHAA accumulation, as it is the case of the lncRNAs targeting different genes related to vitamin C content found in kiwifruit (Deng et al., 2022). Among the other genes that harbor the associated

TABLE 6 Putative candidate genes in the 5.1 Mb region of chromosome 2 containing the 17 single nucleotide polymorphisms (SNPs) associated to DHAA (dehydroascorbic acid) content.

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC111894682	AX-546203383, **AX-546097527**, AX-546203428, **AX-546203454**, AX-546202294, AX-546097531, AX-546202300	**18,547,391**	**18,556,627**	**+**	**GDSL (glycine-asparagine-serine-leucine) esterase/lipase At3g14820 (*Arabidopsis thaliana* GDSL-like lipase/acylhydrolase superfamily)-like**	***GELP***	**Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates**
LOC111894684		18,581,963	18,586,115	–	GDSL esterase/lipase EXL3 (extracellular lipase 3)	*GELP EXL3*	Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates
LOC111894685		18,604,078	18,604,788	+	LOC111894685	LOC111894685	Uncharacterized protein-coding
LOC111892004	**AX-546195535**	**18,907,246**	**18,910,005**	**+**	**GDSL esterase/lipase At5g42170 (*A. thaliana* SGNH (serine-glycine-asparagine-histidine) hydrolase-type esterase superfamily)**	***GELP***	**Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates**
LOC111892005		18,998,678	19,001,618	+	GDSL esterase/lipase At5g42170 (*A. thaliana* SGNH hydrolase-type esterase superfamily)	*GELP*	Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates
LOC111892006		19,138,697	19,140,501	+	LOC111892006	LOC111892006	Uncharacterized long non-coding RNA (lncRNA)
LOC111917458		19,341,283	19,342,294	–	LOC111917458	LOC111917458	Uncharacterized protein-coding
LOC111917457	AX-546196703	19,416,122	19,426,404	–	GDSL esterase/lipase EXL3	*GELP EXL3*	Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates
LOC111917459		19,590,644	19,590,761	+	5S ribosomal RNA (ribonucleic acid)	*5S rRNA*	Enhancement of protein synthesis by stabilizing ribosome structure
LOC111917455		19,610,262	19,616,504	–	GDSL esterase/lipase EXL3	*GELP EXL3*	Catalysis of acyltransfer or hydrolase reactions with lipid and non-lipid substrates
LOC111917456		19,666,560	19,667,356	–	LOC111917456	LOC111917456	Uncharacterized protein-coding
LOC111917454	**AX-546191713**	**19,764,319**	**19,768,646**	**–**	**FREE1 (FYVE (Fab1 (1-phosphatidylinositol 3-phosphate 5-kinase), YOTB, Vac1 (vacuolar transport protein), and EEA1 (early endosome antigen 1)-domain protein required for endosomal sorting) 1**	***FREE1***	**Endosomal sorting complex required for transport (ESCRT) component regulating multivesicular body (MVB) protein sorting, intra-luminal vesicles formation and ubiquitin-dependent protein degradation**

(Continues)

TABLE 6 (Continued)

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC111917451	**AX-546192899**, AX-546192891	**19,772,200**	**19,776,020**	**–**	**FREE1**	***FREE1***	**ESCRT component regulating MVB protein sorting, intra-luminal vesicles formation and ubiquitin-dependent protein degradation**
LOC111917453	AX-546192855, AX-546191621	19,776,750	19,779,968	–	LOC111917453	LOC111917453	Uncharacterized protein-coding
LOC111917450	AX-546192833, AX-546192818, AX-546191589, AX-546191573, AX-546080448, AX-546191538, AX-546192770	19,858,104	19,867,974	–	ABC (ATP (adenosine triphosphate)-binding cassette) transporter B family member 13	*ABCB13*	Auxin efflux transmembrane transporter that is a member of the multidrug resistance P-glycoprotein (MDR/PGP) subfamily of ABC transporters
LOC111916092		20,146,243	20,147,217	+	LOC111916092	LOC111916092	Uncharacterized protein-coding
LOC111916094		20,158,748	20,158,880	+	Small nucleolar RNA104	*snoRNA104*	rRNA processing and maturation
LOC111916081		20,339,463	20,340,294	+	Auxin-responsive protein SAUR (small auxin up-regulated RNA) 50	*SAUR50*	Type 2C phosphatases activity inhibition and cell expansion promotion
LOC111916091		20,608,411	20,611,033	+	LOC111916091	LOC111916091	Uncharacterized protein-coding
LOC111916090		20,617,214	20,620,060	+	LOC111916090	LOC111916090	Uncharacterized protein-coding
LOC111916089		20,620,463	20,621,497	+	LOC111916089	LOC111916089	Uncharacterized protein-coding
LOC111916088		20,639,474	20,647,569	–	LOC111916088	LOC111916088	Uncharacterized lncRNA
LOC111916086		20,658,193	20,660,252	+	LOC111916086	LOC111916086	Uncharacterized protein-coding
LOC128132278		20,669,279	20,681,175	+	LOC128132278	LOC128132278	Uncharacterized protein-coding
LOC128132009		20,682,450	20,683,167	+	LOC128132009	LOC128132009	Uncharacterized lncRNA
LOC111916083	AX-546206772, AX-546207933	20,719,893	20,724,733	–	ATP sulfurylase 2	*APS2*	Catalysis of the first step of the sulfate assimilation pathway
LOC111916082		20,730,886	20,739,630	+	5′−3′ exoribonuclease 3	*XRN3*	Suppression of post-transcriptional gene silencing and processing of pre-rRNAs
LOC111916085		20,792,423	20,823,413	+	UTP (uridine triphosphate):RNA uridylyltransferase 1	*URT1*	Uridylation of mRNAs to reduce accumulation of oligo(A)-tailed mRNAs, repair desadenylated mRNA ends and prevent the biogenesis of siRNAs (small interfering RNAs)

(Continues)

TABLE 6 (Continued)

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC111916084		20,850,280	20,850,687	–	LOC111916084	LOC111916084	Uncharacterized protein-coding
LOC111916093		20,855,711	20,855,817	–	Small nucleolar RNA R71	*snoRNA R71*	rRNA processing and maturation
LOC111920731		21,335,071	21,340,839	+	Cytokinin riboside 5′-monophosphate phosphoribohydrolase LOG7	*LOG7*	Phosphoribohydrolase that converts inactive cytokinin nucleotides to the biologically active free-base forms
LOC111920722		21,351,445	21,359,082	–	Endoglucanase 2	*EGL2*	Cell wall remodeling
LOC111920723		21,459,462	21,465,181	–	LOC111920723	LOC111920723	Uncharacterized protein-coding
LOC111920732		21,492,738	21,494,083	–	Flavonol 3-sulfotransferase	*FST*	O-sulfation of position 3 of flavonols
LOC111920733		21,497,994	21,499,517	–	Flavonol 3-sulfotransferase-like	*FST*	O-sulfation of position 3 of flavonols
LOC111920734	**AX-546186085**	**21,500,026**	**21,502,406**	**–**	**Diphthine methyl ester synthase**	***DPH***	**Catalysis of four methylation reactions of the modified target histidine residue in translation elongation factor 2 (EF2), to form an intermediate called diphthine methyl ester**
LOC111920735	**AX-546186113**	**21,560,289**	**21,564,252**	**+**	**HVA22 (*Hordeum vulgare* abscisic acid-induced protein) 22-like protein G**	***HVA22G***	**ABA (abscisic acid)-induced negative regulation of vesicle transport**
LOC111920737		21,592,197	21,593,540	–	Flavonol 3-sulfotransferase	*FST*	O-sulfation of position 3 of flavonols
LOC111920738		21,597,402	21,599,095	–	Flavonol 3-sulfotransferase-like	*FST*	O-sulfation of position 3 of flavonols
LOC111920739		21,599,488	21,601,854	–	Diphthine methyl ester synthase	*DPH*	Catalysis of four methylation reactions of the modified target histidine residue in translation EF2, to form an intermediate called diphthine methyl ester
LOC111920740		21,605,115	21,611,299	–	Transcription factor E2FC (Elongation Factor 2FC)	*E2FC*	Transcriptional repression of E2F-regulated genes in mature differentiated cells
LOC111920741		21,619,154	21,623,776	–	LOC111920741	LOC111920741	Uncharacterized protein-coding
LOC111920742	**AX-546186176**, AX-546186179, **AX-546187526**, **AX-546186184**, **AX-546186188**	**21,625,830**	**21,630,483**	**–**	**WPP (tryptophan-proline-proline) domain-associated protein**	***WPP***	**Regulation of the mitotic activity in roots. It plays a role with HSP (Heat Shock Protein) 70–1 in facilitating WIT (WPP domain-interacting tail-anchored protein) 1 nuclear envelope targeting**
LOC111920743	**AX-546186226**, AX-546187557	**21,671,729**	**21,675,836**	**+**	**LOC111920743**	**LOC111920743**	**Uncharacterized lncRNA**
LOC111920744		21,680,488	21,681,316	+	LOC111920744	LOC111920744	Uncharacterized lncRNA
LOC128132153		21,738,201	21,739,575	+	LOC128132153	LOC128132153	Uncharacterized lncRNA
LOC111920727		21,740,681	21,742,059	+	LOC111920727	LOC111920727	Uncharacterized protein-coding

(Continues)

TABLE 6 (Continued)

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC111920745		21,743,475	21,746,258	+	MDIS (male discoverer) 1-interacting receptor like kinase 2	*MIK2*	It acts as a receptor of SCOOPs (serine-rich endogenous peptides) regulating multiple processes including plant growth, development and stress responses
LOC128132154		21,746,263	21,749,060	–	LOC128132154	LOC128132154	Uncharacterized lncRNA
LOC111920746		21,762,945	21,767,039	–	Non-specific lipid transfer protein GPI (glycosyl-phosphatidylinositol)-anchored 20	*LTPG20*	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
LOC111920728		21,767,368	21,768,192	+	LOC111920728	LOC111920728	Uncharacterized protein-coding
LOC111920747	**AX-546185962**, AX-546058147, AX-546184536, AX-546185932, AX-546185928	**21,825,566**	**21,829,126**	**+**	**Anion transporter 4;2C chloroplastic**	***ANTR4;2C***	**Inorganic phosphate and probable anion transporter**
LOC111920748	AX-546184483	21,835,623	21,836,464	+	LOC111920748	LOC111920748	Uncharacterized lncRNA
LOC111920729		21,992,636	21,993,391	+	LOC111920729	LOC111920729	Uncharacterized protein-coding
LOC111920749		22,292,212	22,294,828	+	LOC111920749	LOC111920749	Uncharacterized protein-coding
LOC128132280		22,490,627	22,491,776	–	Pectinesterase/pectinesterase inhibitor 47	*PPE47*	Modification of cell walls through demethylesterification or inhibition of demethylesterification in cell wall pectin
LOC111920750		22,530,412	22,534,962	+	LOC111920750	LOC111920750	Uncharacterized protein-coding
LOC111920752		22,546,568	22,549,510	+	LOC111920752	LOC111920752	Uncharacterized lncRNA
LOC122196667		22,628,569	22,629,290	+	LOC122196667	LOC122196667	Uncharacterized protein-coding
LOC111917719		22,701,425	22,731,739	+	NADPH (nicotinamide adenine dinucleotide phosphate)-cytochrome P450 reductase 1-like	*NADPHP450R1*	Electron transfer from NADP to cytochrome P450 in microsomes and to heme oxygenase and cytochrome B5
LOC111917708		22,790,982	22,792,590	–	F-box protein (FBP) At2g21930 (*A. thaliana* F-box associated ubiquitination effector family protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917691		22,805,611	22,807,192	–	FBP CPR (constitutive expressor of PR (pathogenesis-related) genes) 1	*FBP CPR1*	Negative regulation of both salicylic acid (SA)-dependent and SA-independent defense signaling
LOC111917627	**AX-546200493**	**22,812,244**	**22,813,737**	**–**	**FBP At4g19940 (*A. thaliana* F-box and associated interaction domains-containing protein)**	***FBP***	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome

(Continues)

TABLE 6 (Continued)

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC111917682		22,830,699	22,831,826	–	FBP At3g52320 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC122196669		22,876,211	22,877,228	–	LOC122196669	LOC122196669	Uncharacterized lncRNA
LOC111917600		22,904,314	22,908,704	+	LOC111917600	LOC111917600	Uncharacterized protein-coding
LOC111887812		22,910,075	22,918,074	+	LOC111887812	LOC111887812	Uncharacterized protein-coding
LOC128132155		22,910,927	22,911,716	–	LOC128132155	LOC128132155	Uncharacterized lncRNA
LOC111917652		22,921,889	22,923,031	–	FBP At1g30790 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917642		22,970,133	22,971,275	–	FBP At1g11270 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917634		22,984,801	22,985,943	–	FBP At1g11270 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917737		23,205,208	23,206,723	–	FBP At4g21240 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917589		23,284,926	23,286,438	–	FBP At1g47790 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917578	AX-546192898	23,342,399	23,343,866	–	FBP At1g11270 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC111917572		23,352,039	23,353,564	+	LOC111917572	LOC111917572	Uncharacterized lncRNA
LOC111917538		23,359,493	23,364,932	+	LOC111917538	LOC111917538	Uncharacterized lncRNA
LOC111917526	AX-546192844, AX-546194047	23,394,448	23,395,853	+	FBP At1g32420 (*A. thaliana* F-box and associated interaction domains-containing protein)	*FBP*	Regulation of protein degradation tagging them for ubiquitination and subsequent degradation by 26S proteasome
LOC122196440		23,395,967	23,397,609	–	LOC122196440	LOC122196440	Uncharacterized lncRNA
LOC111917726		23,410,176	23,412,200	–	LOC111917726	LOC111917726	Uncharacterized lncRNA
LOC111905617		23,454,527	23,458,447	+	LOC111905617	LOC111905617	Uncharacterized protein-coding

(Continues)

TABLE 6 (Continued)

Gene ID	SNP	Start	End	Strand	Gene name (GFF annotation)	Abbreviated name	Function (according to UniProt)
LOC128132010		23,464,349	23,466,565	+	LOC128132010	LOC128132010	Uncharacterized lncRNA
LOC128132281		23,487,588	23,488,052	+	Extensin-1-like	*EXT1*	Structural component which strengthens the primary cell wall
LOC111903668	**AX-546191437**	**23,533,897**	**23,535,387**	**–**	**F-box/kelch-repeat protein At3g23880 (*A. thaliana* F-box and associated interaction domains-containing protein)**	***KFB***	**FBP of Kelch subfamily that regulates protein degradation by targeting specific substrates**
LOC111903669	**AX-546191457, AX-546192691**	**23,554,314**	**23,563,321**	**+**	**LOC111903669**	**LOC111903669**	**Uncharacterized protein-coding**

Note: Significantly associated SNPs with DHAA content and the genes harboring them are shown in bold.

SNPs with DHAA content, different gene products were identified, like GDSL (glycine-asparagine-serine-leucine) esterase/lipases (LOC111894682, LOC111892004), FREE1 proteins (LOC111917454, LOC111917451), a diphthine methyl ester synthase (LOC111920734), the HVA22-like protein G (LOC111920735), a WPP (tryptophan-proline-proline) domain-associated protein (LOC111920742), an anion transporter (LOC111920747), and two F-box proteins (LOC111917627, LOC111903668) (Table 6). Ten more genes that encode F-box proteins were found in the studied region (Table 6). These results are interesting because a gene encoding an F-box protein has been recently proposed as a candidate gene to regulate the ascorbate peroxidase (APX) activity in a GWAS carried out in barley (Thabet et al., 2022). The APX catalyzes the conversion of AA to DHAA (through ascorbate) as part of the reduction of hydrogen peroxide to water (Apel & Hirt, 2004). F-box proteins have been related to antioxidant status in plants in other previous studies. For example, the overexpression of the *TaFBA1* gene that encodes an F-box protein, enhanced the oxidative stress response through a higher APX activity in wheat (S. M. Zhou et al., 2015). Therefore, the genes in the region of interest in the current study encoding F-box proteins are potential candidates for the control of DHAA content in lettuce. Another interesting candidate gene is the one that encodes the pectinesterase/pectinesterase inhibitor (*PPE*) 47 (LOC128132280) (Table 6). Vitamin C biosynthesis has been widely studied in plants, being the D-manose/L-galactose or Smirnoff-Wheeler pathway (Wheeler et al., 1998) the main route for AA synthesis, though there are at least three alternative pathways. *PPEs* have been previously proposed as candidate genes involved in the increase of vitamin C content through the alternative D-galacturonic acid pathway that starts with the degradation of cell wall pectins (Di Matteo et al., 2010; Ruggieri et al., 2015). In our analysis, we found markers associated with DHAA instead of AA, the last being the compound assessed in both studies just mentioned. However, AA and DHAA contents are directly related as they are interconvertible. Plants need to maintain a balance of antioxidant compounds and DHAA is the product of the AA oxidation mediated by APX (among others), and it is recycled to AA by the action of the dehydroascorbate reductase (DHAR) in the ascorbate-glutathione cycle (Apel & Hirt, 2004). Two more candidate genes, encoding an endoglucanase 2 (EGL2) (LOC111920722) and an extensin-1-like (EXT1) protein (LOC128132281), also play a role in the cell wall remodeling, which could ultimately alter vitamin C content (Table 6). There are also 39 out of the 84 candidate genes that have not been characterized yet in lettuce, including 16 lncRNA (Table 6). They may participate in DHAA accumulation, but more efforts in gene annotation are needed to shed light on their potential functions and contribution to DHAA content in lettuce.

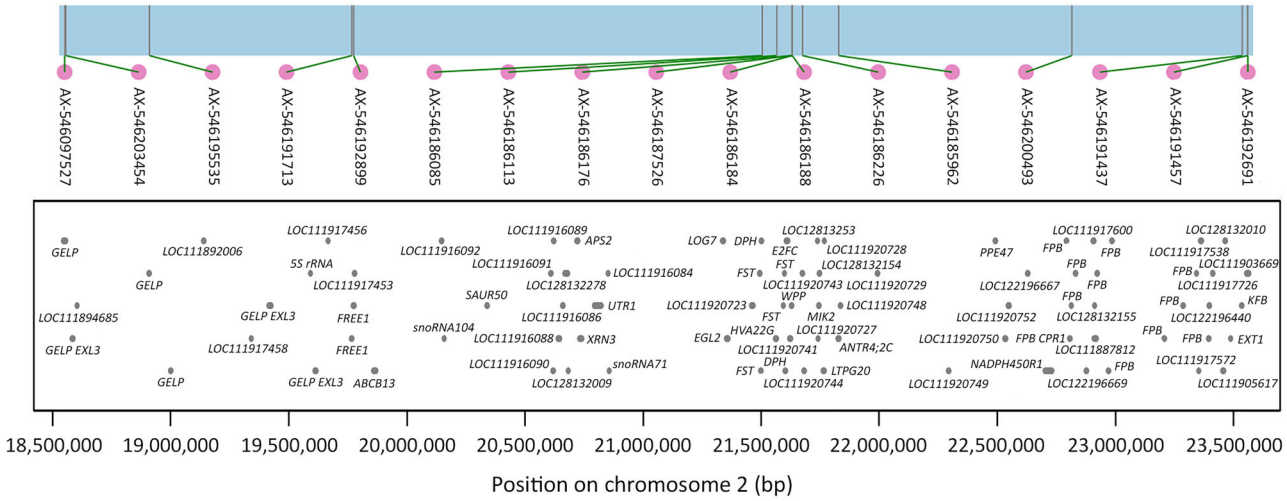


FIGURE 3 Physical map of the 5.1 Mb region containing the 17 single nucleotide polymorphisms (SNPs) significantly associated to dehydroascorbic acid (DHAA) content. Abbreviations of the gene names are described in Table 6.

As far as we know, this is the first GWAS that has found genetic associations with the DHAA content in lettuce. GWAS has been used as a powerful approach in searching for genes or genetic markers associated with different traits in lettuce. However, most studies have assessed resistance to biotic stresses (Lu et al., 2014; Simko et al., 2022, 2023) or agronomic traits like leaf color and morphology, bolting and flowering times, or shelf life and developmental rate (S. Kwon et al., 2013; S. Park et al., 2021; Sthapit Kandel et al., 2020; Wei et al., 2021). In other crops, the regulation of vitamin C content has actually been analyzed by GWAS. Candidate genes related to the common biosynthesis pathways have been reported, such as genes encoding a monodehydroascorbate reductase, a bHLH transcription factor that regulates genes of the D-mannose/L-galactose pathway, and also a PPE like the one identified in our study, what reinforces it as a candidate gene (Berdugo-Cely et al., 2023; Sauvage et al., 2014; Ye et al., 2019). New candidate genes have also been proposed to have an effect on vitamin C content, like the Fas-associated factor 1-like (*FAF1*) and ethylene responsive factor 1 (*ERF1*) genes in tomato (Ruggieri et al., 2014) or genes that encode methionine sulfoxide reductases (*MSR*) in potato (Berdugo-Cely et al., 2023). Except for Sauvage et al. (2014), these studies have evaluated associations exclusively with AA, the most biologically active form of vitamin C, and have not paid attention to DHAA. However, DHAA is essential to maintain the reduced AA pool through its recycling by the action of the DHAR, for instance. In addition, DHAA also presents some biological activity, as mentioned before. Consequently, identifying candidate genes that modulate DHAA content, and not only AA, is essential to understand vitamin C regulation in crops and, in particular, in lettuce. Furthermore, identification of molecular markers associated to DHAA content might allow the use of marker-assisted selection in breeding programs aimed at obtaining vitamin C biofortified lettuce.

AUTHOR CONTRIBUTIONS

Inés Medina-Lozano: Data curation; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Juan Ramón Bertolín**: Data curation; formal analysis; methodology; writing—review and editing. **Jörg Plieske**: Data curation; formal analysis; methodology; writing—review and editing. **Martin Ganal**: Supervision; writing—review and editing. **Heike Gnad**: Resources; writing—review and editing. **Aurora Díaz**: Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

We thank J. A. Aranjuelo and "laboratorio de valoración nutritiva" from CITA for technical support and the staff of SGS Institut Fresenius GmbH TraitGenetics Section for performing the Axiom and Illumina array genotyping. We are also grateful to D. L. Goodchild for reviewing the English language and to the Vegetable Germplasm Bank of Zaragoza (BGHZ-CITA, Spain), the Centre for Genetic Resources (CGN, Wageningen, Netherlands), and Ramiro Arnedo Semillas S.A. for supplying part of the seeds used here. This work was funded by the projects PID2022-138484OR-I00 from the Spanish Ministry of Science and Innovation (MCIN) and State Research Agency (AEI) and AGROALNEXT from the MCIN with funding from European Union (EU) NextGenerationEU: PRTR-C17.I1, and by the European Social Fund from the EU through the Government of Aragon (A12_23R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética").

Inés Medina-Lozano was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the AEI.

CONFLICT OF INTEREST STATEMENT

Jörg Plieske, Heike Gnad, and Martin Ganal are members of the company SGS Institut Fresenius GmbH TraitGenetics Section, which offers lettuce SNP genotyping. The other authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings shown in this study and needed to reproduce the analyses, including the data of the SNP markers, are included in the main manuscript and the supplemental files or are publicly available.

ORCID

Inés Medina-Lozano https://orcid.org/0000-0001-5533-3505
Aurora Díaz https://orcid.org/0000-0001-7297-1699

REFERENCES

Apel, K., & Hirt, H. (2004). Reactive oxygen species: Metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology*, *55*, 373–399. https://doi.org/10.1146/annurev.arplant.55.031903.141701

Berdugo-Cely, J. A., Céron-Lasso, M. D. S., & Yockteng, R. (2023). Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity. *Frontiers in Plant Science*, *13*, 1007104. https://doi.org/10.3389/fpls.2022.1007104

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., & Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science*, *57*(3), 1070–1082. https://doi.org/10.2135/cropsci2016.10.0885

Deng, H., Xia, H., Guo, Y., Liu, X., Lin, L., Wang, J., Xu, K., Lv, X., Hu, R., & Liang, D. (2022). Dynamic changes in ascorbic acid content during fruit development and ripening of *Actinidia latifolia* (an ascorbate-rich fruit crop) and the associated molecular mechanisms. *International Journal of Molecular Sciences*, *23*(10), 5808. https://doi.org/10.3390/ijms23105808

Díaz, A., Martín-Hernández, A. M., Dolcet-Sanjuan, R., Garcés-Claver, A., Álvarez, J. M., Garcia-Mas, J., Picó, B., & Monforte, A. J. (2017). Quantitative trait loci analysis of melon (*Cucumis melo* L.) domestication-related traits. *Theoretical and Applied Genetics*, *130*, 1837–1856. https://doi.org/10.1007/s00122-017-2928-y

Di Matteo, A., Sacco, A., Anacleria, M., Pezzotti, M., Delledonne, M., Ferrarini, A., Frusciante, L., & Barone, A. (2010). The ascorbic acid content of tomato fruits is associated with the expression of genes involved in pectin degradation. *BMC Plant Biology*, *10*, Article 163. https://doi.org/10.1186/1471-2229-10-163

Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., & Yang, T. L. (2021). LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, *22*(4), bbaa227. https://doi.org/10.1093/bib/bbaa227

Doyle, J. J., & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13–15.

Dyląg, A., Smoleń, S., Wisła-Świder, A., Kowalska, I., Sularz, O., Krzemińska, J., Pitala, J., & Koronowicz, A. (2023). Evaluation of the chemical composition and nutritional value of lettuce (*Lactuca sativa* L.) biofortified in hydroponics with iodine in the form of iodoquinolines. *Frontiers in Plant Science*, *14*, 1288773. https://doi.org/10.3389/fpls.2023.1288773

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

FAOSTAT. (2022). *Statistics of the Food and Agriculture Organization of the United Nations*. https://www.fao.org/faostat

Flint-Garcia, S., Feldmann, M. J., Dempewolf, H., Morrell, P. L., & Ross-Ibarra, J. (2023). Diamonds in the not-so-rough: Wild relative diversity hidden in crop genomes. *PLoS Biology*, *21*, e3002235. https://doi.org/10.1371/journal.pbio.3002235

Flint-Garcia, S., Thornsberry, J. M., & Buckler, E. S., IV (2003). Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, *54*(7), 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Granger, M., & Eck, P. (2018). Dietary vitamin C in human health. *Advances in Food and Nutrition Research*, *83*, 281–310. https://doi.org/10.1016/bs.afnr.2017.11.006

Guo, X., Liu, R. H., Fu, X., Sun, X., & Tang, K. (2013). Over-expression of l-galactono-γ-lactone dehydrogenase increases vitamin C, total phenolics and antioxidant activity in lettuce through bio-fortification. *Plant Cell, Tissue and Organ Culture*, *114*, 225–236. https://doi.org/10.1007/s11240-013-0318-y

Hamazaki, K., Kajiya-Kanegae, H., Yamasaki, M., Ebana, K., Yabe, S., Nakagawa, H., & Iwata, H. (2020). Choosing the optimal population for a genome-wide association study: A simulation of whole-genome sequences from rice. *The Plant Genome*, *13*(1), e20005. https://doi.org/10.1002/tpg2.20005

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Kwon, S., Simko, I., Hellier, B., Mou, B., & Hu, J. (2013). Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines. *The Crop Journal*, *1*(1), 25–33. https://doi.org/10.1016/j.cj.2013.07.014

Kwon, S. J., Truco, M. J., & Hu, J. (2012). LSGermOPA, a custom OPA of 384 EST-derived SNPs for high-throughput lettuce (*Lactuca sativa* L.) germplasm fingerprinting. *Molecular Breeding*, *29*(4), 887–901. https://doi.org/10.1007/s11032-011-9623-5

Liu, J., Wang, H., & Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnology Journal*, *13*(3), 319–328. https://doi.org/10.1111/pbi.12336

Liu, K., & Muse, S. V. (2005). PowerMaker: An integrated analysis environment for genetic maker analysis. *Bioinformatics*, *21*(9), 2128–2129. https://doi.org/10.1093/bioinformatics/bti282

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient

genome-wide association studies. *PLOS Genetics*, *12*(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lu, H., Hu, J., & Kwon, S. J. (2014). Association analysis of bacterial leaf spot resistance and SNP markers derived from expressed sequence tags (ESTs) in lettuce (*Lactuca sativa* L.). *Molecular Breeding*, *34*, 997–1006. https://doi.org/10.1007/s11032-014-0092-5

Medina-Lozano, I., Bertolín, J. R., & Díaz, A. (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin C and anthocyanin content. *Food Chemistry*, *359*, 129864. https://doi.org/10.1016/j.foodchem.2021.129864

Medina-Lozano, I., Bertolín, J. R., & Díaz, A. (2024). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Frontiers in Plant Science*, *15*, 3389. https://doi.org/10.3389/fpls.2024.1369658

Medina-Lozano, I., Bertolín, J. R., Zufiaurre, R., & Diaz, A. (2020). Improved UPLC-UV method for the quantification of vitamin C in lettuce varieties (*Lactuca sativa* L.) and crop wild relatives (*Lactuca* spp.). *Journal of Visualized Experiments*, *160*, e61440. https://doi.org/10.3791/61440

Medina-Lozano, I., & Díaz, A. (2021). Nutritional value and phytochemical content of crop landraces and traditional varieties. In A. Elkelish (Ed.), *Landraces —Traditional variety and natural breed* (pp. 95–116). IntechOpen. https://doi.org/10.5772/intechopen.95514

Medina-Lozano, I., & Díaz, A. (2022). Applications of genomic tools in plant breeding: Crop biofortification. *International Journal of Molecular Sciences*, *23*(6), 3086. https://doi.org/10.3390/ijms23063086

Park, J. S., Kang, M. Y., Shim, E. J., Oh, J. H., Seo, K. I., Kim, K. S., Sim, S. C., Chung, S. M., Park, Y., Lee, G. P., Lee, W. S., Kim, M., & Jung, J. K. (2022). Genome-wide core sets of SNP markers and Fluidigm assays for rapid and effective genotypic identification of Korean cultivars of lettuce (*Lactuca sativa* L.). *Horticulture Research*, *9*, uhac119. https://doi.org/10.1093/hr/uhac119

Park, S., Kumar, P., Shi, A., & Mou, B. (2021). Population genetics and genome-wide association studies provide insights into the influence of selective breeding on genetic variation in lettuce. *The Plant Genome*, *14*(2), e20086. https://doi.org/10.1002/tpg2.20086

Peng, H., Zhao, R., Smith, R., & Simko, I. (2022). Phenotypic and genetic analyses of yellow spot malady in lettuce. *Scientia Horticulturae*, *305*, 111389. https://doi.org/10.1016/j.scienta.2022.111389

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, *3*(1), Article 431. https://doi.org/10.1186/2193-1801-3-431

Rstudio Team. (2020). *RStudio: Integrated development environment for R*. RStudio, PBC. http://doi.wiley.com/10.1002/jwmg.232

Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., Milone, M., Cardi, T., Mennella, G., & Barone, A. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biology*, *14*(1), Article 337. https://doi.org/10.1186/s12870-014-0337-9

Ruggieri, V., Sacco, A., Calafiore, R., Frusciante, L., & Barone, A. (2015). Dissecting a QTL into candidate genes highlighted the key role of pectinesterases in regulating the ascorbic acid content in tomato fruit. *The Plant Genome*, *8*(2), plantgenome2014.08.0038. https://doi.org/10.3835/plantgenome2014.08.0038

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., Fernie, A. R., & Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology*, *165*(3), 1120–1132. https://doi.org/10.1104/pp.114.241521

Simko, I. (2023). Dataset on the single nucleotide variation in diversity panel of 500 lettuce accessions genotyped with tunable genotyping-by-sequencing (tGBS) method. *Data in Brief*, *49*, 109419. https://doi.org/10.1016/j.dib.2023.109419

Simko, I., Hasegawa, D. K., Peng, H., & Zhao, R. (2023). Genetic and physiological determinants of lettuce partial resistance to Impatiens necrotic spot virus. *Frontiers in Plant Science*, *14*, 1163683. https://doi.org/10.3389/fpls.2023.1163683

Simko, I., Peng, H., Sthapit Kandel, J., & Zhao, R. (2022). Genome-wide association mapping reveals genomic regions frequently associated with lettuce field resistance to downy mildew. *Theoretical and Applied Genetics*, *135*, 2009–2024. https://doi.org/10.1007/s00122-022-04090-3

Sthapit Kandel, J., Peng, H., Hayes, R. J., Mou, B., & Simko, I. (2020). Genome-wide association mapping reveals loci for shelf life and developmental rate of lettuce. *Theoretical and Applied Genetics*, *133*(6), 1947–1966. https://doi.org/10.1007/s00122-020-03568-2

Sthapit Kandel, J., Sandoya, G. V., Zhou, W., Read, Q. D., Mou, B., & Simko, I. (2022). Identification of quantitative trait loci associated with bacterial leaf spot resistance in baby leaf lettuce. *Plant Disease*, *106*(10), 2583–2590. https://doi.org/10.1094/PDIS-09-21-2087-RE

Stoffel, K., van Leeuwen, H., Kozik, A., Caldwell, D., Ashrafi, H., Cui, X., Tan, X., Hill, T., Reyes-Chin-Wo, S., Truco, M. J., Michelmore, R. W., & Van Deynze, A. (2012). Development and application of a 6.5 million feature Affymetrix Genechip for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics*, *13*, Article 185. https://doi.org/10.1186/1471-2164-13-185

Thabet, S. G., Alomari, D. Z., Börner, A., Brinch-Pedersen, H., & Alqudah, A. M. (2022). Elucidating the genetic architecture controlling antioxidant status and ionic balance in barley under salt stress. *Plant Molecular Biology*, *110*(3), 287–300. https://doi.org/10.1007/s11103-022-01302-8

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*(1), e20077. https://doi.org/10.1002/tpg2.2007

Tripodi, P., Beretta, M., Peltier, D., Kalfas, I., Vasilikiotis, C., Laidet, A., Briand, G., Aichholz, C., Zollinger, T., van Treuren, R., Scaglione, D., & Goritschnig, S. (2023). Development and application of Single Primer Enrichment Technology (SPET) SNP assay for population genomics analysis and candidate gene discovery in lettuce. *Frontiers in Plant Science*, *14*, 1252777. https://doi.org/10.3389/fpls.2023.1252777

UPOV. (2021). *Guidelines for the conduct of tests for distinctness, homogeneity, and stability* (Document UPOV TG/13/11 Rev 2). UPOV.

USDA. (2022). *FoodData central*. https://fdc.nal.usda.gov/

van Treuren, R., van Eekelen, H. D. L. M., Wehrens, R., & de Vos, R. C. H. (2018). Metabolite variation in the lettuce gene pool: Towards healthier crop varieties and food. *Metabolomics*, *14*(11), Article 146. https://doi.org/10.1007/s11306-018-1443-8

Wang, J., & Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Pro-*

teomics and Bioinformatics, *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wei, T., van Treuren, R., Liu, X., Zhang, Z., Chen, J., Liu, Y., Dong, S., Sun, P., Yang, T., Lan, T., Wang, X., Xiong, Z., Liu, Y., Wei, J., Lu, H., Han, S., Chen, J. C., Ni, X., Wang, J., … Liu, H. (2021). Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. *Nature Genetics*, *53*(5), 752–760. https://doi.org/10.1038/s41588-021-00831-0

Weir, B. S., & Hill, W. G. (2002). Estimating F-statistics. *Annual Review of Genetics*, *36*, 721–750. https://doi.org/10.1146/annurev.genet.36

Wheeler, G. L., Jones, M. A., & Smirnoff, N. (1998). The biosynthetic pathway of vitamin C in higher plants. *Nature*, *393*(6683), 365–369. https://doi.org/10.1038/30728

Ye, J., Li, W., Ai, G., Li, C., Liu, G., Chen, W., Wang, B., Wang, W., Lu, Y., Zhang, J., Li, H., Ouyang, B., Zhang, H., Fei, Z., Giovannoni, J. J., Ye, Z., & Zhang, Y. (2019). Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. *PLoS Genetics*, *15*(5), 1008149. https://doi.org/10.1371/journal.pgen.1008149

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, *19*(4), 619–628. https://doi.org/10.1016/j.gpb.2020.10.007

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., Yan, C., Jia, Y., Larkin, R. M., Lavelle, D., Truco, M. J., Chin-Wo, S. R., Michelmore, R. W., & Kuang, H. (2017). RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nature Communications*, *8*(1), 2264. https://doi.org/10.1038/s41467-017-02445-9

Zhang, W., Alseekh, S., Zhu, X., Zhang, Q., Fernie, A. R., Kuang, H., & Wen, W. (2020). Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *The Plant Journal*, *104*(3), 613–630. https://doi.org/10.1111/tpj.14950

Zhou, H., Yu, L., Liu, S., Zhu, A., Yang, Y., Chen, C., Yang, A., Liu, L., & Yu, F. (2023). Transcriptome comparison analyses in UV-B induced AsA accumulation of *Lactuca sativa* L. *BMC Genomics*, *24*(1), Article 61. https://doi.org/10.1186/s12864-023-09133-7

Zhou, S. M., Kong, X. Z., Kang, H. H., Sun, X. D., & Wang, W. (2015). The involvement of wheat F-box protein gene TaFBA1 in the oxidative stress tolerance of plants. *PLoS ONE*, *10*(4), e0122117. https://doi.org/10.1371/journal.pone.0122117

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Medina-Lozano, I., Bertolín, J. R., Plieske, J., Ganal, M., Gnad, H., & Díaz, A. (2024). Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays. *The Plant Genome*, *17*, e20518. https://doi.org/10.1002/tpg2.20518

CHAPTER 2. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://doi.org/10.1002/tpg2.20518.

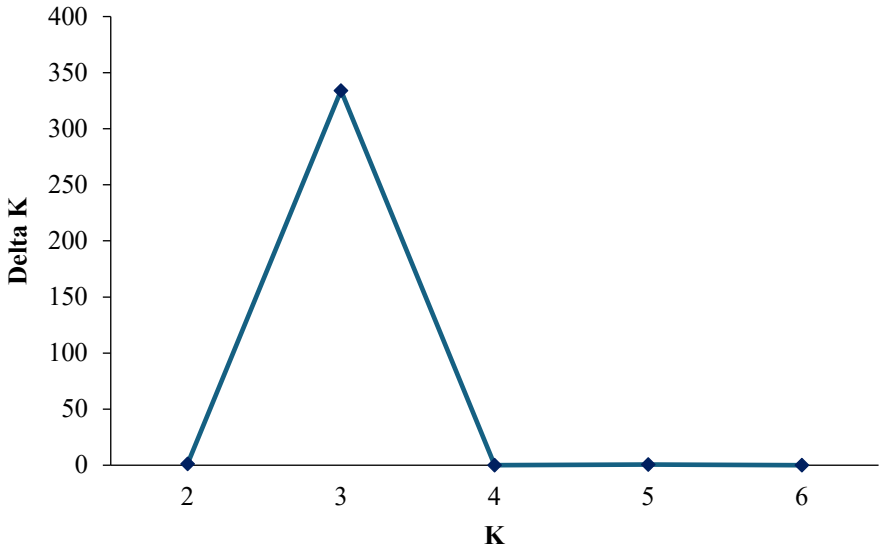


Figure S1. Prediction of optimal number of populations (K) using the delta K (ΔK) method (Evanno et al., 2005).

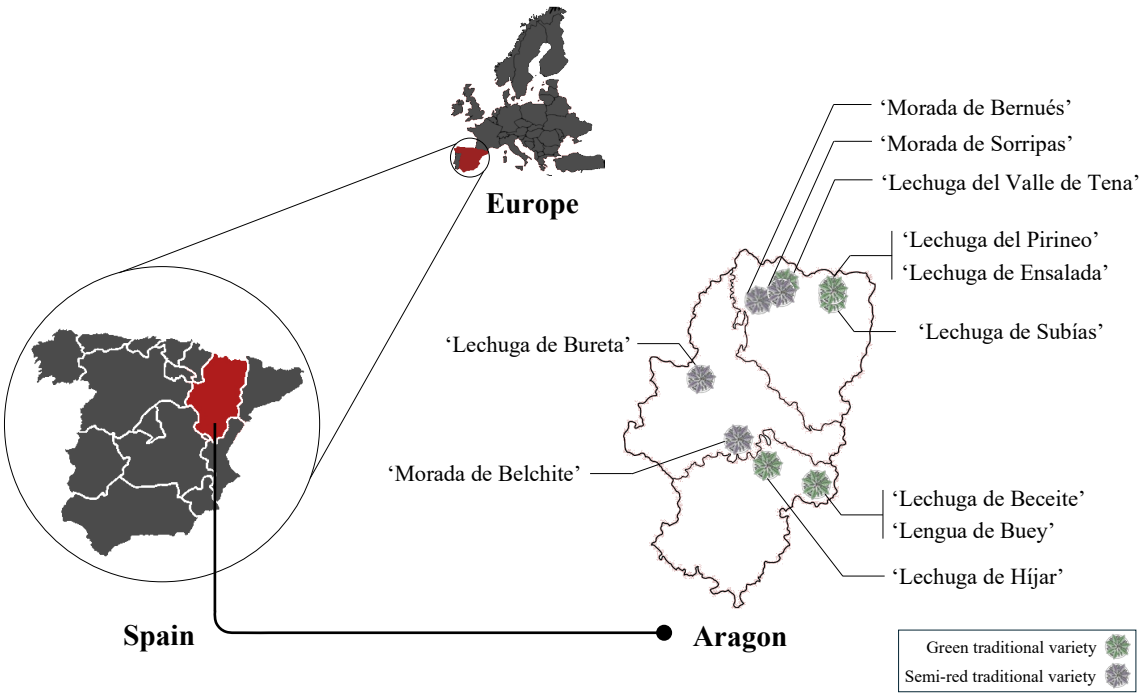


Figure S2. Map of Northeastern Spain with the geographical origin of the lettuce traditional varieties under study.

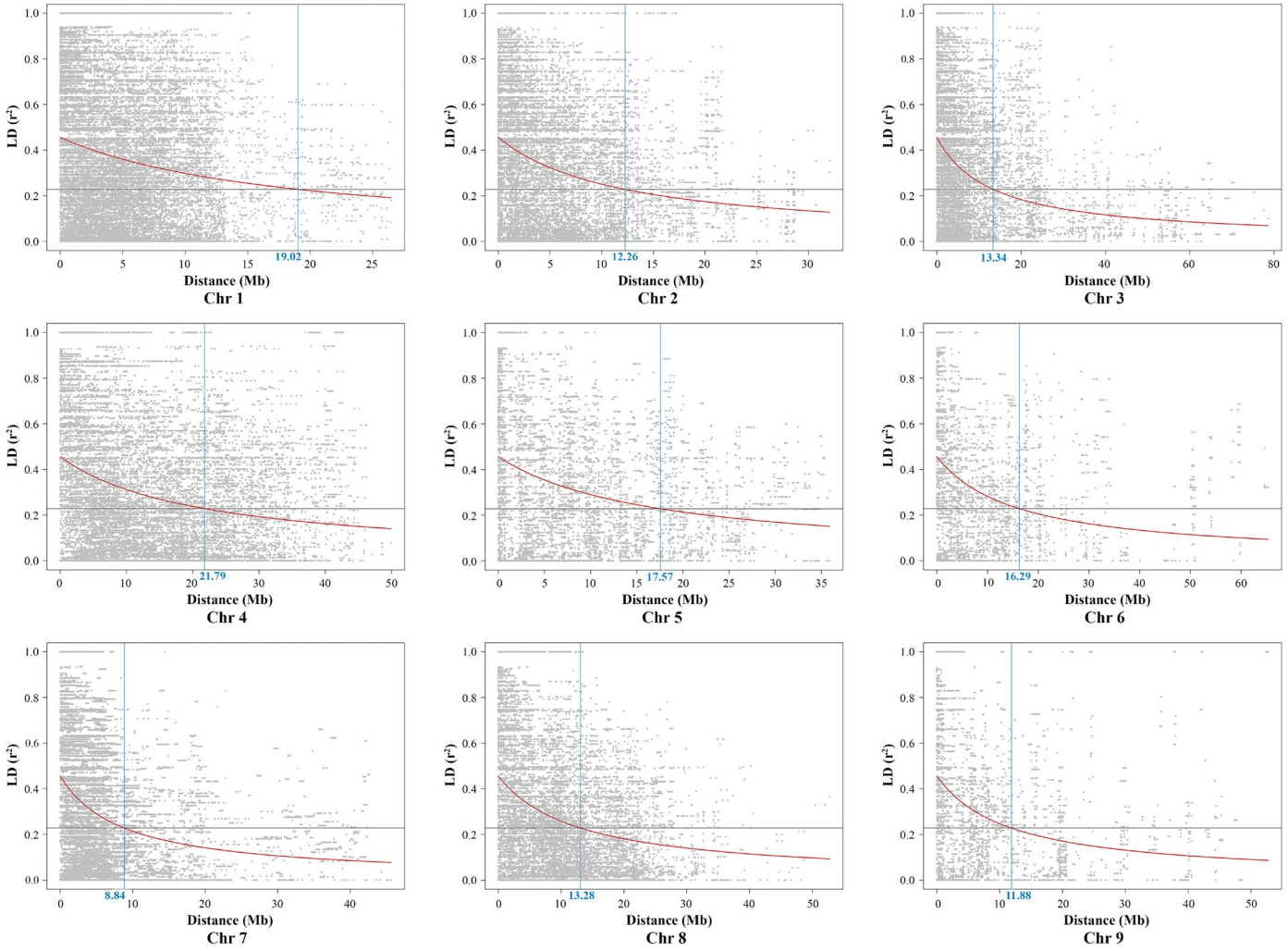


Figure S3. Linkage disequilibrium (LD) decay curve on each chromosome. SNP pairwise correlation coefficients (r^2) were calculated using a windows size of 50 sites around each marker. Chr: chromosome.

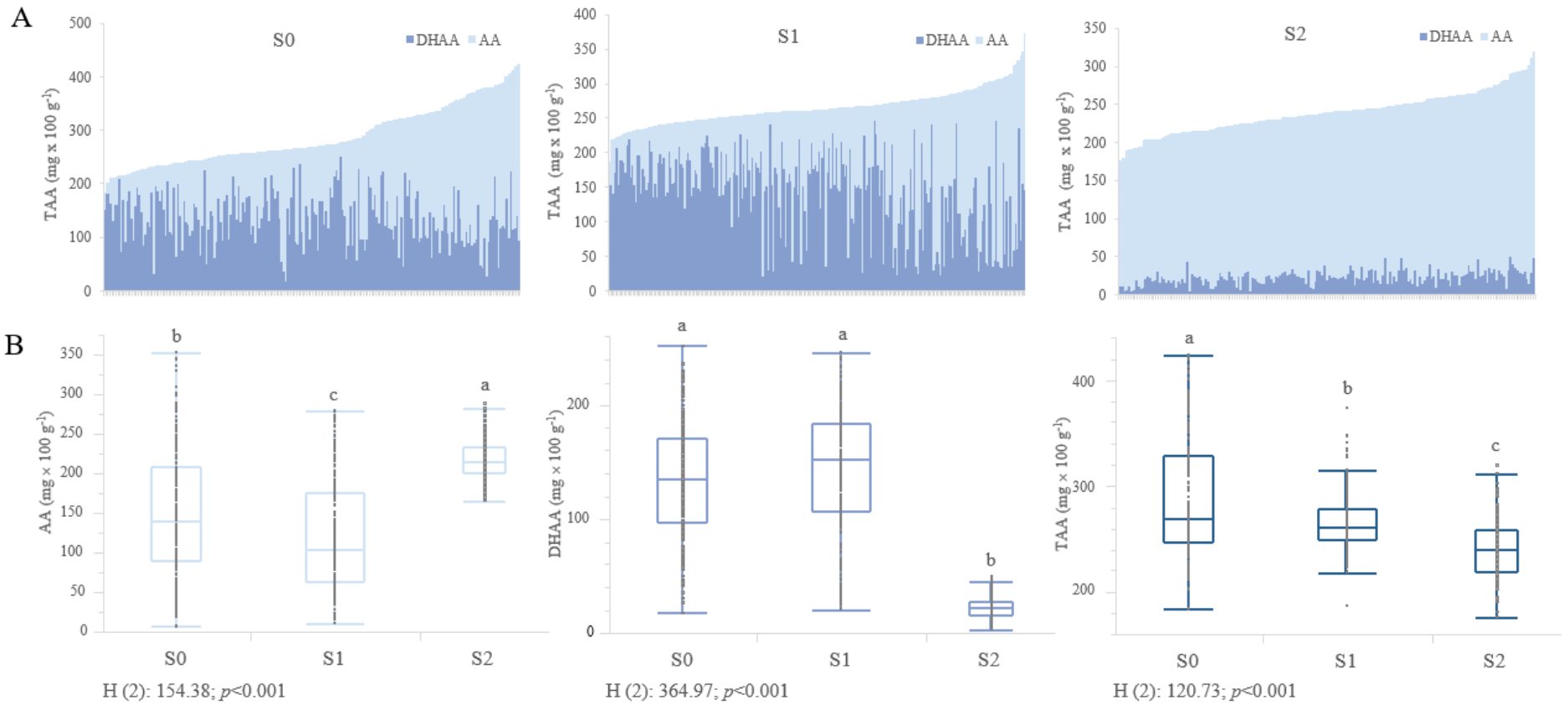


Figure S4. Quantity of total ascorbic acid (TAA), ascorbic acid (AA) and dehydroascorbic acid (DHAA) (A), and results of the Kruskal-Wallis analysis of variance (H statistic (degrees of freedom) and *p* values) and Dunn's test for mean comparison (B) in the original population of the traditional variety 'Lechuga del Pirineo' (205 plants of S0) and two selfing populations derived from the plant with the highest content in TAA (239 S1 and 179 S2 plants). Different letters indicate significant differences among populations at Bonferroni corrected α=0.017.

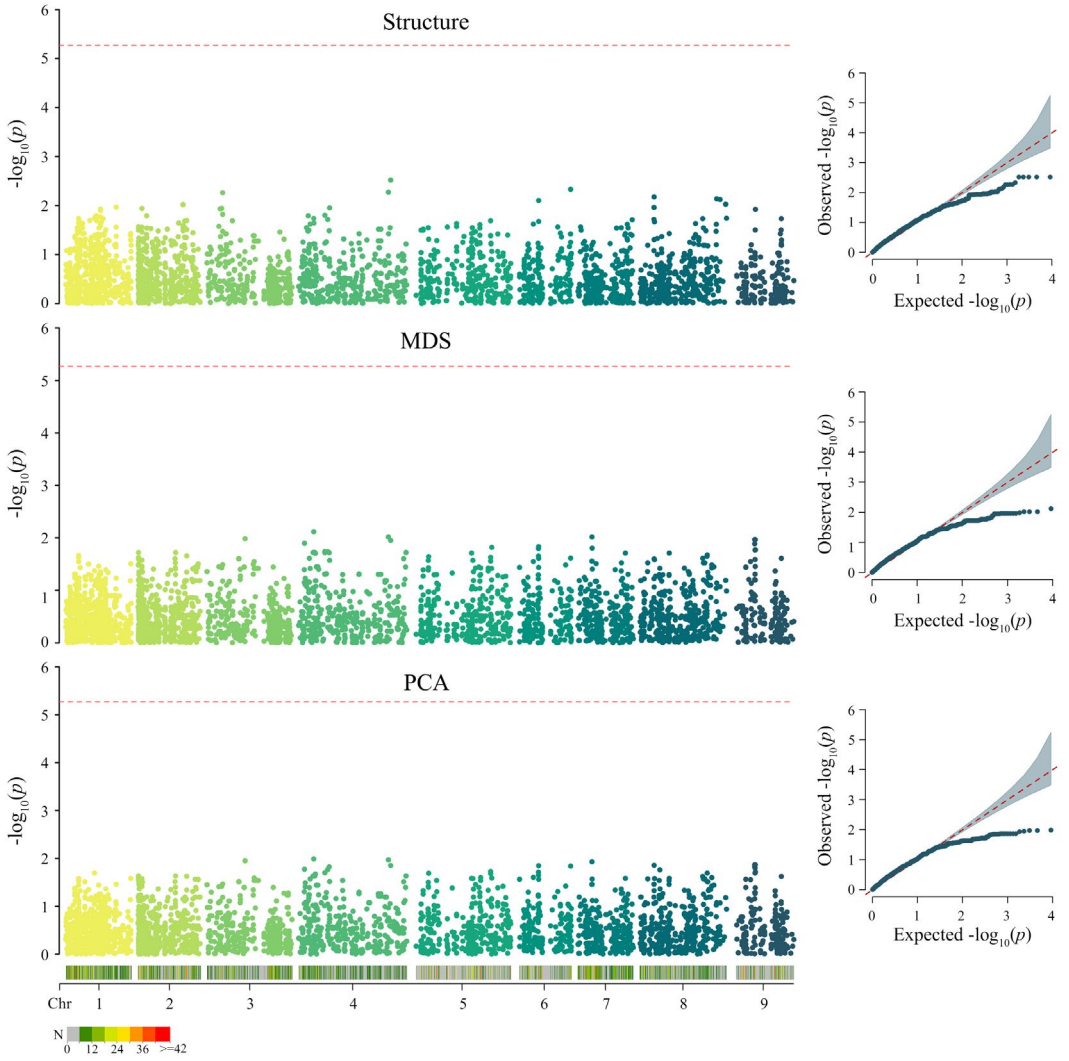


Figure S5. Genome-wide association study (GWAS) for ascorbic acid (AA) content in lettuce samples (21 varieties and 205 plants of 'Lechuga del Pirineo' S0) genotyped with 9,242 SNPs. Manhattan and QQ plots using the Fixed and random model Circulating Probability Unification (FarmCPU) considering both kinship and the structure of the samples measured with three methods: subpopulation membership coefficients (Q) obtained with STRUCTURE software using the linkage model and K = 3, MDS (Multidimensional Scaling) analysis, and PCA (Principal Component Analysis). Statistical significance threshold is shown with the horizontal line ($-\log_{10}(0.05/9{,}242) = 5.27$). Chr: chromosome; N: marker density.

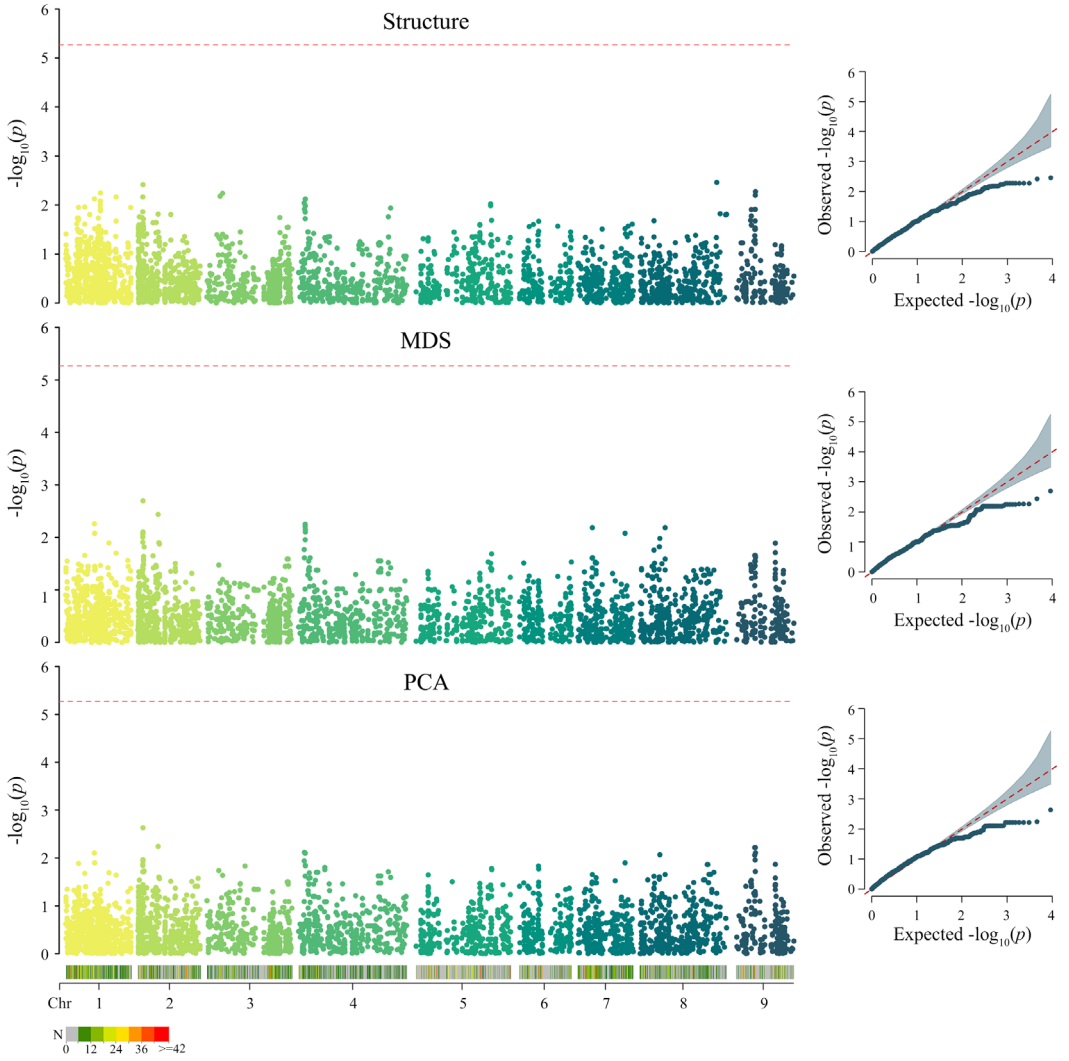


Figure S6. Genome-wide association study (GWAS) for total ascorbic acid (TAA) content in lettuce samples (21 varieties and 205 plants of 'Lechuga del Pirineo' S0) genotyped with 9,242 SNPs. Manhattan and QQ plots using the Fixed and random model Circulating Probability Unification (FarmCPU) considering both kinship and the structure of the samples measured with three methods: subpopulation membership coefficients (Q) obtained with STRUCTURE software using the linkage model and K = 3, MDS (Multidimensional Scaling) analysis, and PCA (Principal Component Analysis). Statistical significance threshold is shown with the horizontal line ($-\log_{10}(0.05/9{,}242) = 5.27$). Chr: chromosome; N: marker density.

Table S1. Description of the 9,242 SNPs used in the GWAS including the alleles, position in the chromosome and *p*-values with the three approaches to assess the population structure, PCA (Principal Component Analysis), MDS (Multidimensional Scaling) and STRUCTURE software for the three traits analyzed, dehydroascorbic acid (DHAA), ascorbic acid (AA), and total ascorbic acid (TAA). Genotypes for all the samples originally included (21 varieties, S0, S1, and S2) are shown.

Available at: https://doi.org/10.1002/tpg2.20518

Table S2. Description of the 13,026 SNPs used in the genetic diversity study including the alleles, position in the chromosome and the genotypes for the 21 lettuce varieties included.

Available at: https://doi.org/10.1002/tpg2.20518

Table S3. Dehydroascorbic acid (DHAA), ascorbic acid (AA) and total ascorbic acid (TAA) (mg × 100 g^{-1}) contents in a diversity panel of 21 lettuce varieties and in three populations of 'Lechuga del Pirineo' (S0, S1 and S2).

Available at: https://doi.org/10.1002/tpg2.20518

CHAPTER 3

Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.)

Medina-Lozano, I, Bertolín, JR, Díaz, A (2024a). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15, 3389. https://doi.org/10.3389/fpls.2024.1369658.

CHAPTER 3

Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.)

Justification

Drought is one of the most threatening environmental factors in the current scenario of climate change, so it is of vital importance to understand the plant response mechanisms with the long-term goal of developing more resilient crops. Nevertheless, water deficit has barely been studied in lettuce, so little is known about the participation of certain antioxidant metabolites in the plant response beyond their antioxidant activity to mitigate drought induced-oxidative stress. Therefore, we have evaluated the changes in the levels of two antioxidant compounds, vitamin C and anthocyanins, in response to drought stress in different lettuce-related accessions (commercial varieties and wild relatives), including members from the groups with the lowest and the highest contents in these compounds according to the classification established in Chapter 1. Previous studies in some other crops have revealed an increase in anthocyanin content in plants subject to drought though this had not been studied in lettuce until now. Regarding the vitamin C, studies conducted in lettuce and other crops did not show such a consistent pattern as in the case of anthocyanins, so we were interested in evaluating it to get a deeper understanding of its behaviour under drought stress conditions.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 March 2024
DOI 10.3389/fpls.2024.1369658

OPEN ACCESS

EDITED BY
Luigi Lucini,
Catholic University of the Sacred Heart, Italy

REVIEWED BY
Leilei Zhang,
Catholic University of the Sacred Heart, Italy
Monica Yorlady Alzate Zuluaga,
Free University of Bozen-Bolzano, Italy
Hajar Salehi,
Catholic University of the Sacred Heart, Italy

*CORRESPONDENCE
Aurora Díaz
adiazb@cita-aragon.es;
adiazb@unizar.es

RECEIVED 18 January 2024
ACCEPTED 22 February 2024
PUBLISHED 15 March 2024

CITATION
Medina-Lozano I, Bertolín JR and Díaz A (2024) Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15:1369658. doi: 10.3389/fpls.2024.1369658

Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.)

Inés Medina-Lozano[1,2], Juan Ramón Bertolín[2,3] and Aurora Díaz[1,2]*

[1]Department of Plant Science, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain, [2]AgriFood Institute of Aragon – IA2, CITA-Universidad de Zaragoza, Zaragoza, Spain, [3]Department of Animal Science, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain

Introduction: Lettuce production and quality could be seriously affected by the increasingly limited water resources.

Methods: The effect of drought on the content of two antioxidant compounds, vitamin C and anthocyanins, in five cultivated lettuces and two wild relatives was assessed for 2 years.

Results and discusion: In leaf samples, *Lactuca* wild species generally had a higher content of total vitamin C than the cultivated lettuces. In contrast, the commercial varieties usually contained more total anthocyanins than the wild species. Total vitamin C decreased with the drought stress in all accessions, commercial varieties, and lettuce wild relatives, with this tendency being consistent and reproducible across the 2 years. These differences were significant in the case of the green commercial varieties 'Winter Crop' (in 2020/2021) and 'Dolomiti G12' (in 2021/2022) and very significant in the red commercial variety 'Red Sails' (in 2020/2021). However, the only group in which the effect of drought was either significant or very significant in both years was the wild species, *Lactuca homblei* and *Lactuca dregeana*, and in the latter also in both tissues (leaf and stem) analyzed. Water stress resulted in an increase of the total anthocyanin content in the leaves from all the accessions, both red commercial varieties and wild relatives, in both years. The most significant enrichment and the only one being either significant or very significant in both years was observed in one of the wild relatives assayed (*L. homblei*). Stems (*L. dregeana*) contained more anthocyanins than leaves under control conditions, and it was exactly the opposite under drought. Changes in anthocyanins in the two tissues in response to drought stress were in opposite directions, increasing in leaves and decreasing in stems. This could suggest a translocation of anthocyanins as a first quick mechanism to cope with a severe lack of water. In

conclusion, anthocyanins (unlike vitamin C) could play a role in the mechanisms deployed by the plant to tolerate drought stress. The wild species with a robust significant enrichment in anthocyanins as a response to drought (*L. homblei*) is a promising plant material to breed more resilient lettuces.

KEYWORDS

lettuce, crop wild relatives, antioxidants, ascorbic acid, UPLC-UV, abiotic stress, water deficit, resilience

1 Introduction

At present, there is a consensus within the scientific community about the planet being immersed in a climate change scenario. There are many consequences of global warming caused or, at least, accelerated by human activity. Among them, one of the most worrying for agriculture consists of the more frequent, longer, and more severe droughts, especially in the Mediterranean Basin (IPCC 2021). This will likely result in economic costs due to crop yield losses, which, in turn, could lead to an increase in food prices. Because of their sessile nature, plants have deployed sophisticated and interconnected mechanisms to survive a wide range of environmental threats (Catalá et al., 2012), and the challenge for scientists lies in uncovering and enhancing them.

Lettuce (*Lactuca sativa* L.) is one of the major leafy vegetable crops worldwide (FAOSTAT, 2021). Over the years, lettuce varieties have been improved through breeding, mainly to increase production and to introduce resistance to diseases. Actually, most of the QTL (quantitative trait loci) or genes mapped in the lettuce genome are responsible for resistance to diseases and pests (Simko et al., 2021). This has resulted in commercial varieties that are, in most cases, nutritionally poor, especially when compared to other leafy vegetables, like chard or spinach (Haytowitz et al., 2015), and not very resilient due to its low tolerance to some adverse environmental conditions, such as drought. In general, crop wild relatives (CWR) have been widely used to increase the crop resistance to biotic stresses (Dempewolf et al., 2017). In contrast, CWR have barely been considered to improve the quality of the crops, even if sometimes they can be richer than the commercial varieties in certain nutrients, like vitamin C, as it happens to be the case in some *Lactuca* spp. (van Treuren et al., 2018; Medina-Lozano et al., 2021). In the case of lettuce, CWR have not been either exploited to enhance the tolerance to abiotic stresses, like drought.

Water deficit is known to cause the accumulation of antioxidants in some crops, though in lettuce, this has barely been studied yet. That is why this research is aimed at assessing the effect of drought stress on the content of two phytochemicals with high antioxidant capacity, vitamin C and anthocyanins, in both cultivated lettuces and wild relatives. Both types of compounds have health-promoting properties mainly due to their antioxidant activity (Cásedas et al., 2020; Xu et al., 2022) and have been reported to participate in the plant defense against biotic stresses and in mechanisms of tolerance to abiotic stresses (Gould, 2004; Locato et al., 2013 and references herein). Vitamin C or total ascorbic acid (TAA) is made up of ascorbic acid (AA) and its oxidation product, dehydroascorbic acid (DHAA) (Medina-Lozano et al., 2020). Vitamin C is an indicator of the quality of many fruits and vegetables that must be incorporated in the diet as it is an essential micronutrient for humans (Carr and Frei, 1999). Anthocyanins are phenolic compounds, specifically flavonoids (Medina-Lozano and Díaz, 2021), responsible for the red color of the leaves of some lettuce varieties. In lettuce, most research on anthocyanins have focused on studying the effect of environmental factors, such as temperature, photoperiod, and wavelength of the artificial light supplied. The content of anthocyanins seems to raise generally in plants when they are grown at low temperatures and with blue light (Li and Ahammed, 2023), which has also been confirmed in lettuce (Chon et al., 2012). Regardless of the environmental effects, the contents of vitamin C and anthocyanins have an important genetic component, with great differences between accessions. Lettuce wild relatives have been described as the richest in vitamin C, followed by the traditional varieties and then the commercial varieties, with exactly the inverse order in the case of the anthocyanins (Medina-Lozano et al., 2021). In other crops, like vine, red grapes accumulate anthocyanins under drought conditions due to a greater and earlier expression of genes of the biosynthesis pathway (Castellarin et al., 2007a; Castellarin et al., 2007b). In fact, in lettuce, the addition of some elicitors, which are phytohormones that participate in a multitude of mechanisms of resistance and tolerance to biotic and abiotic stresses, respectively, increases the content of certain compounds, such as vitamin C (in response to jasmonic acid and arachidonic acid) and some polyphenols (in response to abscisic acid, jasmonic acid, and arachidonic acid) (Złotek et al., 2014). However, a deeper and more thorough investigation on the subject is needed as a very recent study in lettuce landraces shows differing results, with a contrary effect of drought stress on vitamin C (decrease) and phenolic compounds (increase), in response to the water stress (Ćavar Zeljković et al., 2023).

Our hypothesis is that the plant could increase the amount of certain compounds with antioxidant power when subject to stress,

Abbreviations: MPA, meta-phosphoric acid; DTT, 1,4-Dithiothreitol; EDTA, ethylenediaminetetraacetic acid disodium salt.

in this case, drought, as a tolerance mechanism. That is why in this study, we are interested in assessing the changes in some antioxidant compounds (i.e., vitamin C and anthocyanins) in lettuce commercial varieties and some wild relatives under drought stress. Recently, the changes in the amounts of vitamins and phenolic compounds (among other phytochemicals) in response to saline and water stress have been assessed in six cultivated lettuce accessions (Ćavar Zeljković et al., 2023). We believe that it is important to also include CWR in these types of studies because they are expected to show a higher tolerance to adverse environmental conditions. With domestication, crops have become dependent on guaranteed inputs (water, fertilizers, phytosanitary products, etc.), in contrast to the CWR, which will have to cope with the adverse climatic and phytosanitary conditions by evolving endogenous strategies of tolerance and defense. We believe that the enhancement of lettuce resilience could be of great interest as a basis for undertaking future breeding programs.

2 Materials and methods

2.1 Plant material and water stress assays

Seven *Lactuca* accessions were included in this study (Table 1 and Figure 1): five lettuce commercial varieties, two green ('Dolomiti G12' and 'Winter Crop') and three red ('Lollo Rosso', 'Red Sails', and 'Romired'), and two wild relatives (*Lactuca dregeana* DC. and *Lactuca homblei* De Wild). Plants were cultivated for 2 weeks in a growing chamber at 25°C with an average relative humidity of 50% and a short-day photoperiod (10 h light/14 h darkness). Then, they were transplanted to pots (30 × 25 cm and 11.7 L volume) containing a mix of black and blonde peat (1:1) supplemented with fertilizer in a greenhouse at Agrifood Research and Technology Centre of Aragon (CITA, Zaragoza, Spain). Maximum temperature was set at 35°C, relative humidity was set at 40%, and no supplemental lighting was supplied to avoid the

TABLE 1 Description of the plant material used in the water stress study, commercial lettuce varieties, and wild relatives (*Lactuca* spp.).

Name	Species	Group	Leaf color	Source[b]	Accession number
'Dolomiti G12'	*Lactuca sativa* L.	Commercial variety	Green	Ramiro Arnedo Semillas S.A.	–
'Winter Crop'	*Lactuca sativa* L.	Commercial variety	Green	CGN	CGN05853
'Lollo Rosso'	*Lactuca sativa* L.	Commercial variety	Red[a]	CGN	CGN09385
'Red Sails'	*Lactuca sativa* L.	Commercial variety	Red[a]	CGN	CGN19014
'Romired'	*Lactuca sativa* L.	Commercial variety	Red[a]	CGN	CGN24713
Lactuca dregeana	*Lactuca dregeana* DC.	Wild crop relative	Dark green (red stems)	BGHZ	BGHZ3670
Lactuca homblei	*Lactuca homblei* De Wild	Wild crop relative	Green (red nerves)	BGHZ/CGN	BGHZ5322/CGN11322

[a]Semi-red under our experimental conditions.
[b]BGHZ: Vegetable Germplasm Bank of Zaragoza (Spain); CGN: Centre for Genetic Resources (Wageningen, Netherlands).
-, not assigned.

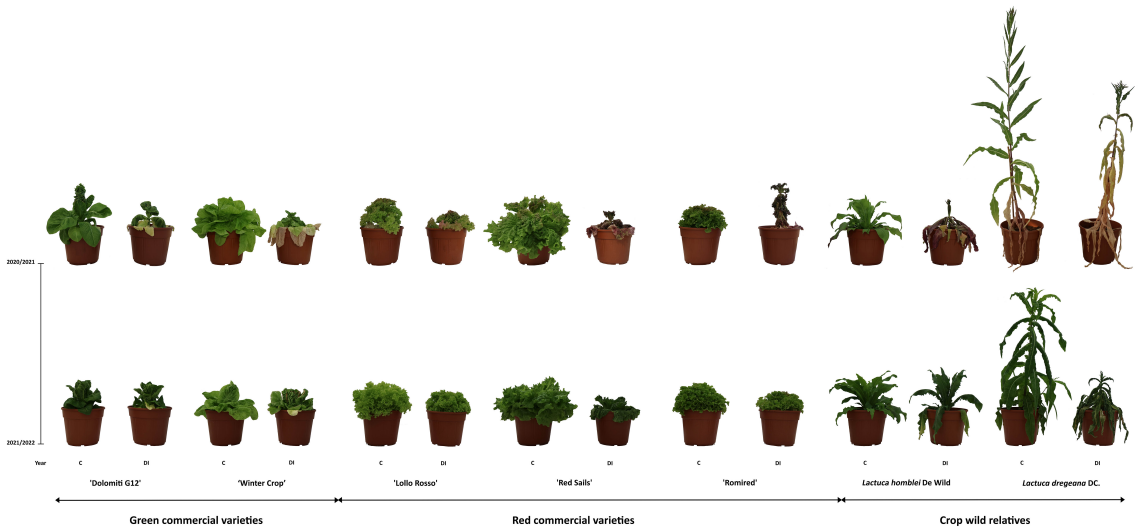


FIGURE 1
Commercial (green and red) lettuce varieties, as well as wild relatives (*Lactuca* spp.), under control (C) and drought stress (DI) experiments carried out during 2020/2021 and 2021/2022.

enhancement of anthocyanin synthesis (natural photoperiod from December to March).

Water stress experiments were repeated two consecutive years, in winters 2020/2021 and 2021/2022. They consisted of two extreme irrigation regimes, control (C) or full irrigation (week 1: 1,350 mL, weeks 2–3: 2,100 mL/each) and deficit irrigation (DI) (weeks 1–3: 0 mL) scheduled 3 weeks before harvesting, which happened to be approximately 3 months after transplanting. More moderate deficit irrigation regimes tested in preliminary experiments carried out in winter 2018/2019 (DI-1: 700 mL and DI-2: 500 mL during only the 2 weeks before harvesting) did not cause any perceivable stress in the plants (data not shown). In both years (2020/2021 and 2021/2022), three plants per accession were assayed in a complete randomized block design, with three biological replicates per block, up to a total of 42 plants.

Two leaves representing the whole plant (i.e., inner and outer leaves in plants that grow forming rosettes or heads), as well as parts of the stem in the case of *L. dregeana* (it grows as a bush), were harvested (Figure 1). All the plant material was immediately frozen using liquid nitrogen and then preserved at −80°C until their lyophilization, as described previously (Medina-Lozano et al., 2020).

2.2 Phytochemical quantification

2.2.1 Extraction and chromatographic determination of vitamin C

As previously described (Medina-Lozano et al., 2020), vitamin C or TAA was quantified as the sum of AA and DHAA in 50 mg of finely powdered lyophilized samples (Table 1). Briefly, AA was extracted with a solution of 8% acetic acid (v/v), 1% MPA (w/v) (both from Sigma-Aldrich, Madrid, Spain), and 1 mM EDTA (Panreac, Barcelona, Spain). After vortexing the samples for 5 s and shaking them for 10 min at 2,000 rpm, they were sonicated for 10 min at room temperature and centrifuged at 4,000 × *g* for 10 min at 4°C. Finally, Extract 1 (E1) was obtained by filtering the supernatant through a 0.22-µm regenerated cellulose filter (Agilent, CA, United States). A 200-µL aliquot of that extract was used to transform DHAA into AA with 200 µL of reducing solution (40 mM DTT with 0.5 M Tris, pH 9.0) (Roche, Madrid, Spain) during 30 min at room temperature in darkness. The reaction was stopped and made stable at acidic pH by adding 200 µL of 0.4 M sulfuric acid (Extract 2, E2).

AA and TAA were quantified by preparing dilutions of E1 and E2, respectively, with ultrapure water (1:4 v:v) and using a liquid chromatographer UPLC H-Class with an HSS T3 column (150 mm × 2.1 mm × 1.8 µm), and an Photodiode Array PDA eλ Detector, controlled by Empower 3 software (Acquity, Waters, Mildford, MA, USA). TAA was composed of AA coming from both sources in E2 (AA present in the extract and AA coming from the reduction of DHAA). DHAA content was easily calculated subtracting AA from TAA. This indirect approach is required because DHAA has a low absorptivity in the UV range of the spectrum, which makes its quantification difficult. The flow rate consisted of 0.3 mL min^{-1} of 2% methanol (ChemLab, Zedelgem, Belgium) and 98% ultrapure water, pH 2.0, acidified with formic acid (Supelco-Sigma-Aldrich) in isocratic mode. A volume of 5 µL of each extract kept at 5°C was injected and run through the column at 30°C for 3 min. AA was identified and quantified using an external calibration curve built with the commercial standard L-ascorbic acid (≥99.9% purity, Sigma-Aldrich) and the photodiode array detector at 245 nm. The analytical method is described in more detail in Medina-Lozano et al. (2020).

2.2.2 Extraction and chromatographic determination of anthocyanins

The extraction of anthocyanins was carried out with samples coming from all the accessions described in Table 1 (green-leaf varieties included) following the method described by Assefa et al. (2019) and adapted by Medina-Lozano et al. (2021) in an environment with low intensity light to prevent anthocyanin degradation. Succinctly, a double extraction with 40 mg of fine powder of lyophilized plant material and 5 mL of methanol: ultrapure water:formic acid solution (50:44:6 v:v:v) was carried out. After vortexing the samples for 5 s and shaking them for 20 min at 2,000 rpm, they were sonicated for 20 min at room temperature and centrifuged at 4,000 × *g* for 10 min at 4°C. Finally, the mixed supernatants from both extractions were passed through a 0.22-µm polytetrafluoroethylene (PTFE) filter (Agilent).

For the chromatographic determination of the anthocyanins, the method described by Fernández-Barbero et al. (2019) with slight modifications (Medina-Lozano et al., 2021) was followed. In brief, the separation of the specific anthocyanins was carried out by liquid chromatography in an Acquity UPLC BEH C18 column (50 mm × 2.1 mm × 1.7 µm, Waters). Methanol (A) and ultrapure water pH 2.0 acidified with formic acid (C) were used as mobile phases at a flow rate of 0.3 mL min^{-1} in gradient mode of A and C (Supplementary Table S1). A volume of 3 µL of each extract kept at 10°C was injected and run through the column at 30°C for 20 min, with chromatograms acquired at 520 nm.

The patterns obtained in a previous work in which these *Lactuca* accessions together with many others were characterized in terms of anthocyanin content (Medina-Lozano et al., 2021) allowed us to identify the same three anthocyanins found before: cyanidin 3-*O*-(6'-*O*-malonylglucoside), cyanidin 3-(6"-acetylglucoside), and peonidin 3-*O*-glucoside. Their quantification was performed building the respective calibration curves, with ≥96% purity Kuromanin chloride or cyanidin 3-*O*-glucoside chloride (from 0.1 to 50 µg mL^{-1}) for the two cyanidin derivates and with ≥95% purity peonidin 3-*O*-glucoside chloride (0.01 to 1 µg mL^{-1}) for the peonidin-type anthocyanin, both purchased from Extrasynthese (Genay, France).

2.3 Method validation

The chromatographic methods were validated in both cases, vitamin C and anthocyanin determination, by calculating different analytical parameters [selectivity, sensitivity, limit of detection (LOD), limit of quantification (LOQ), linearity of the calibration curve tested by the coefficient of determination (R^2), repeatability and intermediate precision expressed as coefficients of variation (CV, %), and recovery (Rec, %) as previously described (Bertolín et al., 2018; Medina-Lozano et al., 2021)].

10.3389/fpls.2024.1369658

2.4 Data analysis

Characteristics in terms of the content of vitamin C (TAA, AA, and DHAA) and of both specific [cyanidin 3-*O*-(6'-*O*-malonylglucoside), cyanidin 3-(6"-acetylglucoside), and peonidin 3-*O*-glucoside] and total anthocyanins by treatment (C *vs.* DI) within accessions were assessed by descriptive statistics in 2020/2021 and 2021/2022 (Table 2; Supplementary Tables S2, S3). One-way analysis of variance (ANOVA) was also employed to compare the means of all the accessions within each irrigation regime (C and DI) for every year (2020/2021 and 2021/2022). The means of all pairs of accessions were compared using the unpaired *t*-test. Non-normally distributed data were transformed {$[TAA]^2$ for DI in 2021/2022 and ln (*anthocyanins*) for C in 2020/2021 and 2021/2022, and for DI in 2021/2022}. Equal variances could be assumed in all cases. The number of biological replicates was 3 ($n = 3$).

The effect of the irrigation regimes assayed (C *vs.* DI) on TAA content in each of the five commercial lettuce varieties and the two wild relatives (*Lactuca* spp.) was studied by unpaired *t*-test ($\alpha = 0.05$) in 2020/2021 and 2021/2022. For the content of anthocyanins, only the three red-leaf commercial varieties plus the two wild species were included as anthocyanins were absent in the two green-leaf commercial varieties, as expected. For the branching wild species (*L. dregeana*), the analyses were carried out with the leaves and also with the stems. Data were transformed to achieve normality only in two cases ($1/[TAA]^{10}$ in 2020/2021 and $[TAA]^{10}$ in 2021/2022) for the red-leaf commercial variety 'Romired'. Equal variances could always be assumed (homoscedasticity was fulfilled).

All the described statistical analyses were conducted with the software JMP v5.1.2 for Windows (SAS Institute Inc., Cary, NC).

3 Results

3.1 Quantification of antioxidant compounds and their changes in response to drought stress

3.1.1 Total vitamin C

In 2020/2021, the mean TAA content in leaves in C conditions ranged between 225.80 and 453.85 mg 100 g^{-1} DW (Table 2; Figure 2A), in a red-leaf commercial variety ('Lollo Rosso') and in a wild relative (*L. homblei*), respectively. In 2021/2022, the results were very similar, with slight movements in the ranking positions with, again, 'Lollo Rosso' being the poorest accession in TAA content in leaves (221.14 mg 100 g^{-1} DW) and *L. homblei* being the richest (345.53 mg 100 g^{-1} DW) (Table 2; Figure 2A). In both years, the differences among accessions were highly significant ($p = 0.00005$ and $p = 0.00003$, respectively).

Under drought stress (DI), the results were again reproducible between years as the accession with the lowest TAA content in leaves was *L. dregeana* (124.87 mg 100 g^{-1} DW in 2020/2021 and

TABLE 2 Total ascorbic acid (TAA) and total anthocyanin content in commercial lettuce varieties and some wild relatives (*Lactuca* spp.) under control (C) and drought stress (DI) conditions in a 2-year experiment.

			2020/2021				2021/2022			
			TAA (mg 100 g^{-1} DW)		Anthocyanins (mg 100 g^{-1} DW)		TAA (mg 100 g^{-1} DW)		Anthocyanins (mg 100 g^{-1} DW)	
Name	Group	Tissue	C	DI	C	DI	C	DI	C	DI
'Dolomiti G12'	Green commercial variety	Leaf	251.73 ± 39.06^{cd}	195.60 ± 11.48^{ab}	ND	ND	245.02 ± 7.42^{cd}	231.17 ± 3.65^{ab}	ND	ND
'Winter Crop'		Leaf	322.58 ± 50.82^{b}	188.79 ± 17.72 ^{ab}	ND	ND	248.03 ± 21.7^{b}	239.73 ± 26.89^{ab}	ND	ND
'Lollo Rosso'	Red commercial variety	Leaf	225.80 ± 21.35^{d}	206.65 ± 12.01^{a}	70.27 ± 43.44^{ab}	74.41 ± 6.17^{cd}	221.14 ± 11.96^{d}	204.49 ± 18.07^{a}	3.01 ± 1.94^{b}	4.32 ± 2.88^{c}
'Red Sails'		Leaf	298.34 ± 11.39^{bc}	195.32 ± 33.65^{ab}	260.8 ± 179.35^{a}	356.9 ± 63.9^{a}	302.43 ± 39.28^{bc}	251.5 ± 24.22^{ab}	15.21 ± 6.96^{a}	26.55 ± 8.75^{a}
'Romired'		Leaf	253.95 ± 65.63^{cd}	213.95 ± 7.31^{a}	147.02 ± 68.19^{a}	180.09 ± 29.63^{bc}	237.04 ± 12.06^{cd}	206.14 ± 47.77^{a}	31.67 ± 5.54^{a}	35.57 ± 12.6^{a}
L. homblei	Wild crop relative	Leaf	453.85 ± 25.56^{a}	224.09 ± 59.18^{a}	32.41 ± 16.59^{bc}	232.29 ± 122.05^{b}	345.53 ± 20.25^{a}	285.14 ± 20.94^{a}	3.8 ± 0.52^{b}	7.02 ± 0.55^{bc}
L. dregeana		Leaf	355.70 ± 8.55^{b}	124.87 ± 65.72^{b}	10.7 ± 7.98^{c}	21.23 ± 7.59^{d}	300.99 ± 13.88^{b}	187.01 ± 28.76^{b}	4.06 ± 3.24^{b}	11.22 ± 3.44^{b}
		Stem	72.84 ± 6.13	33.36 ± 5.24	16.97 ± 8.07	9.82 ± 3.86	126.83 ± 11.79	62.63 ± 24.74	12.18 ± 4.25	7.93 ± 2.73

Different letters indicate significant differences among accessions within each treatment (C, DI) at $p < 0.05$ for leaf samples.
ND: not detected.

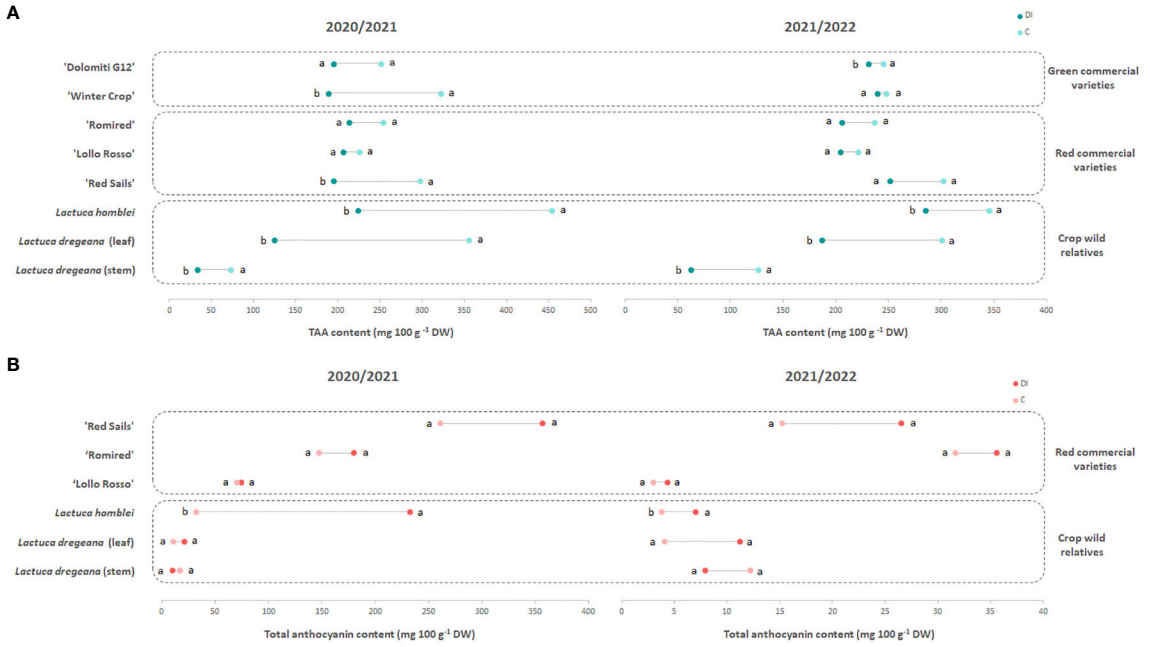


FIGURE 2
Dumbbell plots of the **(A)** TAA (total ascorbic acid) and **(B)** total anthocyanin contents of green and red commercial lettuce varieties and wild relatives (*Lactuca* spp.) under different irrigation regimes, C (for control) or full irrigation (week 1: 1,350 mL, weeks 2–3: 2,100 mL/each) and DI (for deficit irrigation) (weeks 1–3: 0 mL) in 2 consecutive years (2020/2021 and 2021/2022). Different letters show significant differences ($p < 0.05$) between the two irrigation regimes assayed each year.

187.01 mg 100 g^{-1} DW in 2021/2022) and the one with the highest was *L. homblei* (224.09 mg 100 g^{-1} DW in 2020/2021 and 285.14 mg 100 g^{-1} DW in 2021/2022), the latter being also the richest under C conditions (Table 2; Figure 2A), as mentioned above. Even though differences were not significant ($p = 0.1275$) and very significant ($p < 0.01$), respectively, in both years, *t*-test rendered groups with significant differences when paired means were compared (Table 2).

In all cases, in both years and irrigation regimes, the samples with the lowest TAA concentrations were the stem tissues collected from the wild species *L. dregeana* (Table 2; Figure 2A), which was the only plant with bushy growing habits among the studied (Figure 1). Those values were also lower than the TAA content of the leaf samples from the same accession, in total agreement with what has been described before (Medina-Lozano et al., 2021). Considering all data (from both years and the two irrigation regimes) within each accession, AA was the most abundant form of vitamin C in most cases, only in 'Lollo Rosso' and 'Dolomiti G12' was DHAA either always or in half of the cases the predominant compound, respectively (Supplementary Tables S2, S3).

The accessions under study were chosen to gather a wide range of variability, from those with a high content in TAA, that is, the CWR (*L. dregeana* and *L. homblei*), to those poor in TAA, that is, the cultivated varieties, both green ('Dolomiti G12' and 'Winter Crop') and red ('Lollo Rosso', 'Red Sails' and 'Romired'), according to the classification from Medina-Lozano et al. (2021).

TAA content decreased with the drought stress (DI) in all accessions, commercial varieties (both green and red), and CWR, with that tendency being consistent and reproducible across the 2 years in which the experiments were carried out (Table 2). In most cases, the drop in TAA quantity could be observed in both forms of vitamin C, AA and DHAA, and only in a few cases did the reduction happen in only one of them, mainly in AA (Supplementary Tables S2, S3). These differences in TAA content in leaves between treatments (C *vs.* DI) were significant in the case of the green commercial varieties 'Winter Crop' (in 2020/2021; $p = 0.013$) and 'Dolomiti G12' (in 2021/2022; $p = 0.044$) and very significant in the red commercial variety 'Red Sails' (in 2020/2021; $p = 0.007$) and the two CWR, *L. homblei* (in 2020/2021, $p = 0.004$ and in 2021/2022, $p = 0.023$) and *L. dregeana* (in both 2020/2021 and 2021/2022, $p = 0.004$). In *L. dregeana* stems, the differences were also statistically significant (in 2020/2021, $p = 0.001$ and in 2021/2022, $p = 0.046$) (Table 3; Figure 2A).

3.1.2 Total anthocyanins

Anthocyanins were quantified in all the accessions (Table 1; Figure 1), the two green commercial varieties included, which served as negative controls of the method because, as expected, they did not contain any at all (Table 2; Supplementary Tables S2, S3).

In 2020/2021, the average total anthocyanin content in leaves in C conditions ranged between 10.70 and 260.80 mg 100 g^{-1} DW (Table 2; Figure 2B). The accession with the lowest content was a CWR (*L. dregeana*) and the one with the highest value was a red commercial variety ('Red Sails'). A similar tendency was observed for the DI treatment the same year, with *L. dregeana* (21.23 mg 100 g^{-1} DW) and 'Red Sails' (356.90 mg 100 g^{-1} DW) again showing the lowest and the highest content in total anthocyanins, respectively. The differences were very significant ($p = 0.003$) and highly significant ($p = 0.0005$) in C and DI, respectively.

10.3389/fpls.2024.1369658

TABLE 3 Effect of drought stress conditions assayed on the total ascorbic acid (TAA) and total anthocyanin contents in commercial lettuce varieties and some wild relatives (*Lactuca* spp.) in a 2-year experiment.

			TAA				Anthocyanins			
			2020/2021		2021/2022		2020/2021		2021/2022	
Name	Group	Tissue	t Ratio	Prob > |t|[a]	t Ratio	Prob > |t|[a]	t Ratio	Prob > |t|[a]	t Ratio	Prob > |t|[a]
'Dolomiti G12'	Green commercial variety	Leaf	2.388	0.075	2.900	0.044	–	–	–	–
'Winter Crop'		Leaf	4.305	0.013	0.416	0.699	–	–	–	–
'Lollo Rosso'	Red commercial variety	Leaf	1.354	0.247	1.331	0.254	−0.163	0.878	−0.650	0.551
'Red Sails'		Leaf	5.023	0.007	1.912	0.129	−0.874	0.431	−1.756	0.154
'Romired'		Leaf	−1.046[b]	0.355[b]	1.182[b]	0.303[b]	−0.770	0.484	−0.491	0.649
L. homblei	Wild crop relative	Leaf	6.173	0.004	3.591	0.023	−2.811	0.048	−7.345	0.002
L. dregeana		Leaf	6.033	0.004	6.182	0.004	−1.655	0.173	−2.625	0.059
		Stem	8.480	0.001	3.299	0.046	1.384	0.239	1.406	0.254

[a]$\alpha = 0.05$ used to test significant differences in TAA and total anthocyanin contents between control and deficit irrigation conditions.
[b]Data transformed ($1/[TAA]^{10}$ in 2020/2021 and $[TAA]^{10}$ in 2021/2022) to achieve normality.
-, not analysed because no anthocyanins were detected.

In 2021/2022, the values of the mean total anthocyanin content in leaves varied between 3.01 and 31.67 mg 100 g^{-1} DW in 'Lollo Rosso' and 'Romired', respectively, under C conditions, and between 4.32 and 35.57 mg 100 g^{-1} DW again in 'Lollo Rosso' and 'Romired', respectively, under DI conditions (Table 2; Figure 2B). Again, the differences were very significant (p = 0.001) and highly significant (p = 0.0003) in C and DI, respectively. As observed in the case of TAA content, the tendencies were comparable between treatments and years.

Interestingly, quite the opposite to what happened with TAA, in C conditions, the total anthocyanin content was higher in the stem than in the leaf in the wild species *L. dregeana* in both years (Table 2; Figure 2B), in total agreement with previous observations (Medina-Lozano et al., 2021). However, the situation was reversed under drought stress (DI); that is, leaves had more total anthocyanins than stems (in the same plants), as we will discuss in more detail below. The same three specific anthocyanins, cyanidin 3-*O*-(6'-*O*-malonylglucoside), cyanidin 3-(6"-acetylglucoside), and peonidin 3-*O*-glucoside, found previously in a bigger set of accessions, including the ones studied here (Medina-Lozano et al., 2021), were also identified in this occasion (Supplementary Tables S2, S3).

As described before in another bushy lettuce wild relative, *L. squarrosa* (Medina-Lozano et al., 2021), in this work, the cyanidin 3-(6"-acetylglucoside) was present only in the stems but not in the leaves of *L. dregeana* (Supplementary Table S2), under both C and DI conditions in 2020/2021 [cyanidin 3-(6"-acetylglucoside) was not detected at all in 2021/2022 probably due to the low levels of total anthocyanins]. Cyanidin 3-(6"-acetylglucoside) was not found in any of the cultivated varieties included in the previous study, though now we have detected it in two red-leaf commercial varieties, 'Lollo Rosso' and 'Romired' in 2020/2021 (Supplementary Table S2), probably due to the high quantities of total anthocyanins accumulated that year, as other authors have also reported it in some other red lettuce varieties (Wu & Prior, 2005; Viacava et al., 2017). Conversely, peonidin 3-*O*-glucoside, which was only observed in cultivated varieties in that same previous study, has also been detected here in the wild species *L. homblei* (Supplementary Tables S2; S3), but exclusively under DI treatment, as previously described.

The selected accessions covered all the spectrum in terms of anthocyanin content, from those with an absolute lack of them (the green commercial varieties 'Dolomiti G12' and 'Winter Crop') used to validate the method, or with a low content (i.e., the CWR *L. homblei*), to those with a high content (the red commercial varieties 'Red Sails' and 'Romired'), with some accessions containing a medium amount in between (the red commercial variety 'Lollo Rosso' and the CWR *L. dregeana*), according to the groups described by Medina-Lozano et al. (2021).

In the case of anthocyanins, the response to drought stress was also very robust across the 2 years in which the two irrigation regimes (C *vs.* DI) were tested (Table 2; Figure 2B). The water stress conditions resulted in an increase in the total anthocyanin content in the leaf samples from all the accessions, both red-leaf commercial varieties and CWR in both years (as expected, anthocyanins could not be detected in the green-leaf varieties). Interestingly, the increase in total anthocyanin content in response to drought stress was significant and very significant in the case of the CWR *L. homblei* in 2020/2021 (p = 0.048) and 2021/2022 (p = 0.002), respectively (Table 3; Figure 2B).

The only case in which the total anthocyanin quantity was lower in plants under DI regime was in the stem tissue of the wild accession *L. dregeana* (Table 2; Figure 2B). Again, these results were reproducible across the 2 years. As mentioned before, stems contained more anthocyanins than leaves under C conditions, and it was exactly the opposite under DI conditions, where leaves had more anthocyanins than stems, and the changes in anthocyanins in the two tissues in response to drought stress were in inverse directions, increasing in leaves and decreasing in stems (Table 2;

Figure 2B) in the same plant and for the three biological replicates (data not shown).

For the leaf samples, in all cases (all accessions and both years), the quantity of the major anthocyanin, cyanidin 3-*O*-(6′-*O*-malonylglucoside), increased as a consequence of water deprivation (DI). Only in the case of the stem samples did its quantity decrease, as expected because there was a drop in total anthocyanin content (Supplementary Table S2, S3).

4 Discussion

4.1 Antioxidant compound levels and variations under drought stress

4.1.1 Total vitamin C

The results regarding TAA content in leaves under C conditions are in the same order of magnitude and in accordance with those obtained in a previous study in which these same accessions, among many others, were characterized for their quantity of vitamin C, which revealed that, generally, the *Lactuca* wild species had a higher content of TAA than the cultivated lettuces (Medina-Lozano et al., 2021). Something similar was observed by van Treuren et al. (2018) when comparing the AA content in lettuce varieties and some of their wild relatives.

AA and DHAA are easily interconvertible, though AA is the form with the highest antioxidant potential. That could explain why most cases in which DHAA was more abundant happened under C conditions, whereas under DI (drought stress), AA was predominant. Moreover, DHAA content was found higher than AA only in commercial varieties; that is, the lettuce wild relatives were always richer in the compound with more antioxidant capacity, AA (Supplementary Tables S2, S3).

The consistent detrimental effect of the water stress on the TAA content in our study totally agrees with what has been recently found in several leafy vegetables (Park et al., 2023) and in some cultivated lettuces (Ćavar Zeljković et al., 2023), in which drought also caused a decrease in their vitamin C amounts.

4.1.2 Total anthocyanins

In contrast to what was obtained for vitamin C and in total agreement with our previous work (Medina-Lozano et al., 2021), the leaves from commercial lettuce varieties contained more total anthocyanins on average than the wild species assessed under C conditions.

Out of the three different anthocyanins found in this work and in a previous study also from our group (Medina-Lozano et al., 2021), cyanidin 3-*O*-(6'-*O*-malonylglucoside) was always present, and it was the most abundant one, as reported also by other authors (Wu and Prior, 2005; Mulabagal et al., 2010; Becker et al., 2014) whereas cyanidin 3-(6''-acetylglucoside) was the rarest. However, some differences have been observed. For instance, 'Lollo Rosso' contained only cyanidin 3-*O*-(6'-*O*-malonylglucoside) in our previous study and in the second year of this study, while in the first year, it contained all three. Similarly, 'Romired' had only cyanidin 3-*O*-(6'-*O*-malonylglucoside) and peonidin 3-*O*-glucoside when it was previously characterized (Medina-Lozano et al., 2021) and in the second year of this work, while in the first year, the three anthocyanins were identified [though cyanidin 3-(6''-acetylglucoside) was present only in the samples under DI treatment]. Similar cases are also observed in the CWR assessed. In general, when there was a high quantity of total anthocyanins, as it happened in 2020/2021, all three of them were present whereas when it was low, the less abundant ones could either not be detected or not be produced. Thus, what can be deduced from this is that some accessions have the genes to synthesize all those anthocyanins even if they are not produced all the time or even in all the parts of the plant. Actually, there were cases in which one of the minor anthocyanins was only synthesized when the plant was under drought stress, DI in our experiments [i.e., peonidin 3-*O*-glucoside in *L. homblei* in both years, and cyanidin 3-(6''-acetylglucoside) in 'Romired' in 2020/2021; presumably, it would have been detected also in 2021/2022 if the total anthocyanin amount would have been higher] (Supplementary Tables S2, S3). That was actually observed in all the biological repeats (data not shown). These observations, together with the rise in the total amount of anthocyanins in all the tested accessions in response to the drought stress, point out a putative role of these compounds in the mechanisms deployed by the plants to tolerate these adverse conditions.

In most cases, only considering leaf tissue and C conditions, red commercial varieties had a higher content of total anthocyanins than the wild species, in contrast to what happened for vitamin C (Table 2). Both observations totally agree with results found by our group previously (Medina-Lozano et al., 2021). In the case of anthocyanins, that was expectable as red-leaf varieties have been bred to enhance their anthocyanic color because it has a positive effect on their market price (Missio et al., 2018).

The significant increase of anthocyanins in *L. homblei* under water stress observed in both years points out to a potential source of drought tolerance in this wild species that could be easily introduced in the cultivated lettuce as they both belong to the same genus *Lactuca*. It is actually common for plants to synthesize different compounds to actively fight against adverse conditions, for instance, sugars, antifreeze proteins, or heat shock proteins in response to low temperatures, frost, and high temperatures, respectively, among others (Catalá et al., 2012). These results are in total agreement with our hypothesis of anthocyanins playing a role in the response of lettuce-related species to drought stress. In line with this, a recent study in vine has demonstrated the induction of *WRKY40*, a transcription factor that regulates anthocyanin biosynthesis, in the berry skin when the plants were subjected to severe water stress (Carvalho et al., 2023).

Under drought stress, the situation is the opposite in terms of anthocyanin content in the two tissues studied. That is, under C conditions, the stems had a higher amount of anthocyanins than the leaves, whereas under drought stress, the leaves contained more anthocyanins than the stems. These results could suggest a

translocation of these antioxidant compounds as a first quick mechanism to cope with a severe lack of water in a tissue with a high susceptibility to water deprivation such as the leaf (from less to more susceptible tissues to dehydration). In this way, anthocyanins would be located where they are most needed when plant survival is under threat. It seems that there are routes for the transport of anthocyanins at subcellular, cellular, tissue, and organ levels (Zhao & Dixon, 2010 and references herein) though their regulation, specially under stress, has not been properly studied yet. This putative movement of phytochemicals in response to water stress was not observed in the case of TAA, which makes sense as its quantity decreases under DI, which rules out TAA having a role in the drought tolerance response.

The rise in anthocyanins as a consequence of the lack of irrigation reported in this study is in line with the results reported by Ćavar Zeljković et al., (2023) that observed an increase in some phenolic compounds (free phenolic acids and flavonoids) in response to drought in some cultivated lettuces, though the authors did not include anthocyanins among the flavonoids quantified.

The inverted tendencies shown by TAA and total anthocyanin contents in response to the deficit irrigation (Figure 1; Table 2) lined up with our previous observations, that is, the existence of a negative correlation between the content of both types of compounds (Medina-Lozano et al., 2021).

5 Conclusions

Water deprivation caused a drop in vitamin C content in all accessions within all groups and in both tissues—leaf and stem—analyzed. The explanation to this could be that, in a critical moment in which the survival is under threat, the plant could shut down all the processes that are not directly involved in survival and tolerance to drought stress, in this case. Another non-exclusive reason could be that the rise in anthocyanin biosynthesis could cause a reduction in other processes of the plant that compete directly for the same resources. Quite the opposite to what happened to vitamin C, the anthocyanin content increased under water stress in the leaf samples from all accessions within all groups (it only decreased in the stem samples). According to the hypothesis just revealed, it is reasonable to think that anthocyanins could play a role in the mechanisms deployed by the plant to cope with the drought stress though this needs to be verified in future experiments. If so, these would be promising plant materials for breeding more drought-tolerant lettuces that are also rich in some compounds that provide health benefits. Interestingly, a significant enrichment in anthocyanins as a response to the drought stress was observed in one of the wild relatives assayed (*L. homblei*). It is well known that CWR are a source of resistance in the case of biotic stresses. Similarly, it is expected to find higher levels of tolerance to abiotic stresses among the CWR as they are exposed to adverse climatic conditions usually mitigated for the crops by any agricultural system.

The challenge ahead is to decipher the role of anthocyanins in drought tolerance to engineer them and improve the lettuce resilience and, as a positive side effect, its benefits to human health. If so, a disadvantageous situation like drought could be turned into something beneficial in terms of plant fitness and human nutrition. Furthermore, it would also be possible to reduce the volume of irrigation water without paying any production penalty, which has undeniable environmental and economic upsides.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

IML: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. JRB: Data curation, Investigation, Methodology, Writing – review & editing. AD: Data curation, Investigation, Methodology, Writing – review & editing, Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the projects PID2022-138484OR-I00 from the Spanish Ministry of Science and Innovation and State Research Agency (AEI) and LMP164_18 and LMP148_21, both from the Government of Aragón; and by the Operational Programme FEDER Aragón 2014-2020 and 2020-2022, and the European Social Fund from the European Union (A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética"). IM-L was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Spanish State Research Agency (AEI).

Acknowledgments

We thank J. A. Aranjuelo and "laboratorio de valoración nutritiva" from CITA for technical support and D. L. Goodchild for reviewing the English language. We gratefully acknowledge the Vegetable Germplasm Bank of Zaragoza (BGHZ-CITA, Spain), the Centre for Genetic Resources (CGN, Wageningen, Netherlands), and Ramiro Arnedo Semillas S.A. for supplying the seeds used here.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1369658/full#supplementary-material

References

Assefa, A. D., Choi, S., Lee, J.-E., Sung, J.-S., Hur, O.-S., Ro, N.-Y., et al. (2019). Identification and quantification of selected metabolites in differently pigmented leaves of lettuce (*Lactuca sativa* L.) cultivars harvested at mature and bolting stages. *BMC Chem.* 13, 56. doi: 10.1186/s13065-019-0570-2

Becker, C., Klaering, H.-P., Schreiner, M., Kroh, L. W., and Krumbein, A. (2014). Unlike Quercetin Glycosides, Cyanidin Glycoside in Red Leaf Lettuce Responds More Sensitively to Increasing Low Radiation Intensity before than after Head Formation Has Started. *J. Agric. Food Chem.* 62, 6911–6917. doi: 10.1021/jf404782n

Bertolín, J. R., Joy, M., Rufino-Moya, P. J., Lobón, S., and Blanco, M. (2018). Simultaneous determination of carotenoids, tocopherols, retinol and cholesterol in ovine lyophilised samples of milk, meat, and liver and in unprocessed/raw samples of fat. *Food Chem.* 257, 182–188. doi: 10.1016/j.foodchem.2018.02.139

Carr, A. C., and Frei, B. (1999). Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. *Am. J. Clin. Nutr.* 69, 1086–1107. doi: 10.1093/ajcn/69.6.1086

Carvalho, L. C., Ramos, M. J. N., Faísca-Silva, D., Marreiros, P., Fernandes, J. C., Egipto, R., et al. (2023). Modulation of the berry skin transcriptome of cv. Tempranillo induced by water stress levels. *Plants* 12, 1778. doi: 10.3390/plants12091778

Cásedas, G., Les, F., and López, V. (2020). Anthocyanins: plant pigments, food ingredients or therapeutic agents for the CNS? A mini-review focused on clinical trials. *Curr. Pharm. Des.* 26, 1790–1798. doi: 10.2174/1381612826666200127093701

Castellarin, S. D., Matthews, M. A., Di Gaspero, G., and Gambetta, G. A. (2007a). Water deficits accelerate ripening and induce changes in gene expression regulating flavonoid biosynthesis in grape berries. *Planta* 227, 101–112. doi: 10.1007/s00425-007-0598-8

Castellarin, S. D., Pfeiffer, A., Sivilotti, P., Degan, M., Peterlunger, E., and Di Gaspero, G. (2007b). Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ.* 30, 1381–1399. doi: 10.1111/j.1365-3040.2007.01716.x

Catalá, R., Díaz, A., and Salinas, J. (2012). "Molecular responses to extreme temperatures," in *Plant biotechnology and agriculture: Prospects for the 21st century*. Eds. A. Altman and P. M. Hasegawa (San Diego, CA: Elsevier), 287–307.

Ćavar Zeljković, S., Štefelová, N., Hron, K., Doležalová, I., and Tarkowski, P. (2023). Preharvest abiotic stress affects the nutritional value of lettuce. *Agronomy* 13, 398. doi: 10.3390/agronomy13020398

Chon, S.-U., Boo, H.-O., Heo, B.-G., and Gorinstein, S. (2012). Anthocyanin content and the activities of polyphenol oxidase, peroxidase and phenylalanine ammonia-lyase in lettuce cultivars. *Int. J. Food Sci. Nutr.* 63, 45–48. doi: 10.3109/09637486.2011.595704

Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885

FAOSTAT (2021). *Statistics of the Food and Agriculture Organization of the United Nations*.

Fernández-Barbero, G., Pinedo, C., Espada-Bellido, E., Ferreiro-González, M., Carrera, C., Palma, M., et al. (2019). Optimization of ultrasound-assisted extraction of bioactive compounds from jabuticaba (*Myrciaria cauliflora*) fruit through a Box-Behnken experimental design. *Food Sci. Technol.* 39, 1018–1029. doi: 10.1590/fst.16918

Gould, K. S. (2004). Nature's swiss army knife: the diverse protective roles of anthocyanins in leaves. *J. BioMed. Biotechnol.* 2004, 314–320. doi: 10.1155/S1110724304406147

Haytowitz, D. B., Ahuja, J. K. C., Showell, B., Somanchi, M., Nickle, M., and Nguyen, Q. A. (2015). Composition of Foods Raw, Processed, Prepared. *USDA National Nutrient Database for Standard Reference, Release 28*. Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, ARS, USDA. Dataset. doi: 10.15482/USDA.ADC/1324304

Li, Z., and Ahammed, G. J. (2023). Plant stress response and adaptation via anthocyanins: A review. *Plant Stress* 10, 100230. doi: 10.1016/j.stress.2023.100230

Locato, V., Cimini, S., and De Gara, L. (2013). Strategies to increase vitamin C in plants: from plant defense perspective to food biofortification. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00152

Medina-Lozano, I., Bertolín, J. R., and Díaz, A. (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin C and anthocyanin content. *Food Chem.* 359, 129864. doi: 10.1016/j.foodchem.2021.129864

Medina-Lozano, I., Bertolín, J. R., Zufiaurre, R., and Díaz, A. (2020). Improved UPLC-UV method for the quantification of vitamin C in lettuce varieties (*Lactuca sativa* L.) and crop wild relatives (*Lactuca* spp.). *JOVE-J Vis. Exp.* 160, e61440. doi: 10.3791/61440

Medina-Lozano, I., and Díaz, A. (2021). "Nutritional value and phytochemical content of landraces and traditional varieties," in *Landraces - Traditional Variety and Natural Breed* Ed. A, Elkelish (IntechOpen), pp. 95–116.

Missio, J. C., Rivera, A., Figàs, M. R., Casanova, C., Camí, B., Soler, S., et al. (2018). A comparison of landraces vs. Modern varieties of lettuce in organic farming during the winter in the mediterranean area: an approach considering the viewpoints of breeders, consumers, and farmers. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01491

Mulabagal, V., Ngouajio, M., Nair, A., Zhang, Y., Gottumukkala, A. L., and Nair, M. G. (2010). *In vitro* evaluation of red and green lettuce (*Lactuca sativa*) for functional food properties. *Food Chem.* 118, 300–306. doi: 10.1016/j.foodchem.2009.04.119

Park, T., Fischer, S., Lambert, C., Hilger, T., Jordan, I., and Cadisch, G. (2023). Combined effects of drought and soil fertility on the synthesis of vitamins in green leafy vegetables. *Agriculture* 13, 984. doi: 10.3390/agriculture13050984

Simko, I., Jia, M., Venkatesh, J., Kang, B.-C., Weng, Y., Barcaccia, G., et al. (2021). Genomics and marker-assisted improvement of vegetable crops. *CRC Crit. Rev. Plant Sci.* 40, 303–365. doi: 10.1080/07352689.2021.1941605

van Treuren, R., van Eekelen, H. D. L. M., Wehrens, R., and de Vos, R. C. H. (2018). Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. *Metabolomics* 14, 146. doi: 10.1007/s11306-018-1443-8

Viacava, G. E., Roura, S. I., Berrueta, L. A., Iriondo, C., Gallo, B., and Alonso-Salces, R. M. (2017). Characterization of phenolic compounds in green and red oak-leaf lettuce cultivars by UHPLC-DAD-ESI-QToF/MS using MS^E scan mode. *J. Mass Spectrometry* 52, 873–902. doi: 10.1002/jms.4021

Wu, X., and Prior, R. L. (2005). Identification and characterization of anthocyanins by high-performance liquid chromatography–electrospray ionization–tandem mass spectrometry in common foods in the United States: vegetables, nuts, and grains. *J. Agric. Food Chem.* 53, 3101–3113. doi: 10.1021/jf0478861

Xu, K., Peng, R., Zou, Y., Jiang, X., Sun, Q., and Song, C. (2022). Vitamin C intake and multiple health outcomes: an umbrella review of systematic reviews and meta-analyses. *Int. J. Food Sci. Nutr.* 73, 588–599. doi: 10.1080/09637486.2022.2048359

Zhao, J., and Dixon, R. A. (2010). The 'ins' and 'outs' of flavonoid transport. *Trends Plant Sci.* 15, 72–80. doi: 10.1016/j.tplants.2009.11.006

Złotek, U., Świeca, M., and Jakubczyk, A. (2014). Effect of abiotic elicitation on main health-promoting compounds, antioxidant activity and commercial quality of butter lettuce (*Lactuca sativa* L.). *Food Chem.* 148, 253–260. doi: 10.1016/j.foodchem.2013.10.031

CHAPTER 3. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1369658/full#supplementary-material.

Table S1. Gradient conditions applied to identify and quantify by UPLC-UV the anthocyanins present in commercial lettuce varieties and some wild relatives (*Lactuca* spp.) under control and drought stress experiments carried out during 2020/2021 and 2021/2022.

Time	Flow (mL/min)	%A	%C	Curve
Initial	0.3	15	85	Lineal
3.30	0.3	20	80	Lineal
3.86	0.3	30	70	Lineal
5.05	0.3	40	60	Lineal
5.35	0.3	55	45	Lineal
5.64	0.3	60	40	Lineal
5.95	0.3	95	5	Lineal
7.50	0.3	95	5	Lineal
12.00	0.3	95	5	Lineal
12.30	0.3	15	85	Lineal
17.00	0.3	15	85	Lineal

Table S2. Vitamin C and anthocyanin content in commercial lettuce varieties and some wild relatives (*Lactuca* spp.) under control (C) and deficit irrigation (DI) experiments carried out during 2020/2021.

			Vitamin C (mg 100 g^{-1} DW)				Anthocyanins (mg 100 g^{-1} DW)[a]					
			AA		DHAA		Cyanidin 3-*O*-(6'-*O*-malonylglucoside)		Cyanidin 3-(6''-acetylglucoside)		Peonidin 3-*O*-glucoside	
Samples	**Group**	**Tissue**	**C**	**DI[b]**	**C**	**DI[b]**	**C**	**DI[b]**	**C**	**DI[b]**	**C**	**DI[b]**
'Dolomiti G12'	Green commercial variety	Leaf	103.40 ± 63.95	150.84 ± 36.4	148.33 ± 30.05	44.76 ± 28.15	ND	ND	ND	ND	ND	ND
'Winter Crop'		Leaf	97.47 ± 30.01	135.04 ± 23.38	225.11 ± 26.66	53.76 ± 13.56	ND	ND	ND	ND	ND	ND
'Lollo Rosso'	Red commercial variety	Leaf	77.72 ± 41.54	90.73 ± 14.72	148.08 ± 21.05	115.91 ± 10.4	67.67 ± 41.03	71.01 ± 5.89	0.57 ± 0.57	1.13 ± 0.41	2.03 ± 1.92	2.27 ± 1.32
'Red Sails'		Leaf	130.94 ± 19.24	156.90 ± 27.72	167.40 ± 30.05	38.42 ± 12.17	254.69 ± 174.77	344.64 ± 61.99	ND	ND	6.11 ± 4.97	12.26 ± 1.94
'Romired'		Leaf	168.34 ± 17.58	167.25 ± 26.75	85.61 ± 49.11	46.70 ± 19.92	142.72 ± 65.27	172.34 ± 28.41	ND	2.59 ± 0.29	4.30 ± 2.94	5.17 ± 1.07
L. homblei	Wild crop relative	Leaf	299.77 ± 81.66	115.98 ± 15.31	154.08 ± 99.87	108.11 ± 73.36	32.41 ± 16.59	224.52 ± 118.55	ND	ND	ND	7.77 ± 3.51
L. dregeana		Leaf	322.35 ± 12.62	108.22 ± 55.09	33.35 ± 20.03	16.65 ± 10.62	10.70 ± 7.98	21.23 ± 7.59	ND	ND	ND	ND
		Stem	54.23 ± 12.60	23.32 ± 1.33	18.61 ± 11.24	10.04 ± 4.81	16.24 ± 7.84	9.18 ± 3.60	0.73 ± 0.31	0.64 ± 0.27	ND	ND

[a]ND: not detected
[b]DI: deficit irrigation (weeks 1-3: 0 mL)

Table S3. Vitamin C and anthocyanin content in commercial lettuce varieties and some wild relatives (*Lactuca* spp.) under control (C) and deficit irrigation (DI) experiments carried out during 2021/2022.

			Vitamin C (mg 100 g^{-1} DW)				Anthocyanins (mg 100 g^{-1} DW)[ab]			
			AA		DHAA		Cyanidin 3-*O*-(6'-*O*-malonylglucoside)		Peonidin 3-*O*-glucoside	
Samples	**Group**	**Tissue**	**C**	**DI[c]**	**C**	**DI[c]**	**C**	**DI[c]**	**C**	**DI[c]**
'Dolomiti G12'	Green commercial variety	Leaf	206.95 ± 9.40	112.31 ± 25.38	38.07 ± 7.73	118.86 ± 22.16	ND	ND	ND	ND
'Winter Crop'		Leaf	199.89 ± 19.51	193.25 ± 25.28	48.14 ± 3.10	46.48 ± 3.36	ND	ND	ND	ND
'Lollo Rosso'	Red commercial variety	Leaf	49.24 ± 8.75	39.97 ± 7.18	171.90 ± 18.50	164.52 ± 10.98	3.01 ± 1.94	4.32 ± 2.88	ND	ND
'Red Sails'		Leaf	180.44 ± 139.31	213.49 ± 17.04	121.99 ± 100.05	38.01 ± 7.68	15.22 ± 6.96	26.55 ± 8.75	ND	ND
'Romired'		Leaf	174.54 ± 16.22	158.00 ± 31.24	62.50 ± 6.65	47.14 ± 16.66	30.56 ± 4.76	34.61 ± 12.31	1.10 ± 0.82	0.96 ± 0.46
L. homblei	Wild crop relative	Leaf	301.61 ± 35.31	252 ± 19.77	43.93 ± 15.34	33.06 ± 5.34	3.80 ± 0.52	5.33 ± 1.33	ND	1.70 ± 0.87
L. dregeana		Leaf	249.57 ± 9.49	168.81 ± 24.5	51.42 ± 7.95	18.14 ± 5.69	4.06 ± 3.24	11.22 ± 3.44	ND	ND
		Stem	104.48 ± 8.61	48.71 ± 20.00	22.35 ± 3.18	13.92 ± 6.96	12.18 ± 4.25	7.93 ± 2.73	ND	ND

[a]ND: not detected
[b]Cyanidin 3-(6''-acetylglucoside) was not detected in 2021/2022
[c]DI: deficit irrigation (weeks 1-3: 0 mL)

CHAPTER 4

Transcriptional dissection of the increase in anthocyanin content under drought stress conditions in a lettuce commercial variety and a wild relative and polymorphism mining in differentially expressed genes

Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2023). Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives. *Int. J. Mol. Sci.* 24, 3052. https://doi.org/10.3390/ijms24033052.

Medina-Lozano, I, Grimplet, J, Díaz, A (2025). Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress. *Front. Plant Sci*. 15, 1494339. https://doi.org/10.3389/fpls.2024.1494339.

CHAPTER 4

Transcriptional dissection of the increase in anthocyanin content under drought stress conditions in a lettuce commercial variety and a wild relative and polymorphism mining in differentially expressed genes

Justification

Anthocyanins were the compounds whose content increased when lettuce-related accessions were subjected to drought stress, whereas vitamin C content consistently decreased, as shown in the previous chapter. This suggests that anthocyanins (but not vitamin C) could be implied in the response to drought in lettuce and wild relatives. Therefore, a transcriptomic analysis was carried out using one representative accession of both commercial varieties and wild relatives with the aim of identifying genes related to the increased anthocyanin content under water deficit (second paper of the chapter). To validate the expression profiles of those candidate genes, reliable reference genes (RGs) for the normalization of real-time quantitative PCR (qPCR) data are required. That is why the current chapter also includes a previous paper reporting the selection of the most suitable RGs for the drought stress experiment (among others), based on RNA-seq data.

International Journal of
Molecular Sciences

MDPI

Article

Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives

Inés Medina-Lozano [1,2], María Soledad Arnedo [3], Jérôme Grimplet [1,2] and Aurora Díaz [1,2,*]

[1] Department of Plant Sciences, Agrifood Research and Technology Centre of Aragon (CITA), Avd. Montañana 930, 50059 Zaragoza, Spain
[2] AgriFood Institute of Aragon–IA2, CITA-University of Zaragoza, 50013 Zaragoza, Spain
[3] Ramiro Arnedo S.A. Paraje La Molina 54, Las Norias de Daza, 04716 Almería, Spain
* Correspondence: adiazb@cita-aragon.es

Abstract: Lettuce is a popular vegetable source of bioactive compounds, like anthocyanins, powerful antioxidants present in red and semi-red varieties. Selection of reliable reference genes (RGs) for the normalisation of real-time quantitative PCR (qPCR) data is crucial to obtain accurate gene expression results. Among the genes with totally unrelated biological functions, six candidate RGs (*ADF2*, *CYB5*, *iPGAM*, *SCL13*, *TRXL3-3*, and *VHA-H)* with low variation in expression according to RNA-seq analyses, were selected for future expression studies of anthocyanin-related genes in three different experiments: leaf colour comparison (green vs. red) in commercial varieties; tissue comparison (leaf vs. stem) in a wild relative; and drought stress experiment in commercial and traditional varieties, and a wild relative. Expression profiles of the candidate RGs were obtained by qPCR and their stability was assessed by four different analytical tools, geNorm, NormFinder, BestKeeper, and Delta Ct method, all integrated in RefFinder. All results considered, we recommend *CYB5* to be used as RG for the leaf colour experiment and *TRXL3-3* for the tissue and drought stress ones, as they were the most stable genes in each case. RNA-seq is useful to preselect novel RGs although validation by qPCR is still advisable. These results provide helpful information for gene expression studies in *Lactuca* spp. under the described conditions.

Keywords: BestKeeper; Delta Ct method; drought stress; geNorm; *Lactuca sativa* L.; NormFinder; leaf colour; RefFinder; tissues; transcriptomics

Citation: Medina-Lozano, I.; Arnedo, M.S.; Grimplet, J.; Díaz, A. Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives. *Int. J. Mol. Sci.* **2023**, *24*, 3052. https://doi.org/10.3390/ijms24033052

Academic Editor: Zsófia Bánfalvi

Received: 30 November 2022
Revised: 23 January 2023
Accepted: 26 January 2023
Published: 3 February 2023

1. Introduction

Lettuce (*Lactuca sativa* L.) is one of the most important leafy vegetables worldwide and its popularity is yearly increasing. Moreover, it is a valuable source of bioactive compounds [1], which have beneficial effects on human health. One of them are anthocyanins, water-soluble phenolic compounds with powerful antioxidant properties that help to prevent some human diseases, like cardiovascular and neuronal disorders, cancer or diabetes [2].

In lettuce, anthocyanins are responsible for the red colour, so they are exclusively present in red and semi-red varieties. In some previous studies, transcriptomic analyses have been carried out to compare green and red-leaf varieties [3–5] or to elucidate the accumulation of anthocyanins in lettuce [6,7]. There is a limited number of gene expression studies carried out in tissues different from leaves, like roots and stems, either in lettuce or its wild relatives. Among them, very few are focused on genes involved in the biosynthesis of bioactive compounds, like *LsHPPD* (4-hydroxyphenylpyruvate dioxygenase), a gene encoding for an enzyme from the pathway of vitamin E biosynthesis [8]. In this sense, lettuce wild relatives, e.g., *Lactuca serriola* L., have been used to explore the differential expression of candidate genes responsible for the synthesis of antioxidant compounds [7,9]. Transcriptomic studies to evaluate the response to

different abiotic stresses, e.g., cold [10], heat [11,12], heavy metals [13,14], drought [15,16] and UV radiation [17] have been performed in lettuce. However, before this work, anthocyanin regulation has not been studied either under drought stress or in different tissues in the genus *Lactuca*.

Real-time quantitative PCR (qPCR) is nowadays the most widely used technique for gene expression studies, due to its high specificity and sensitivity, as well as its immediacy in delivering results [18]. Nevertheless, a correct quantification of gene expression relies on an accurate normalisation of qPCR data that minimizes differences resulting from variation among samples and experimental errors. The most common normalisation method is the use of the so-called reference genes (RGs) as internal reaction controls. Ideally, RGs have to exhibit a relatively constant expression regardless of the nature of the samples and the experimental conditions [19].

Traditionally, RGs have been selected among those considered as housekeeping genes, since they encode proteins implied in essential cellular processes and then, they are expected to show stable expression throughout the different tissues and conditions. Some typical examples of housekeeping genes used as RGs in plants are actin (*Act*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), elongation factor 1-α (*eEF1-α*), ubiquitin (*Ubq*), tubulin (*Tub*), and 18S ribosomal RNA (*18SrRNA*) [20–23]. However, the use of conventional housekeeping genes as RGs is not always appropriate. This has been demonstrated in plants by many studies in which these genes show notable expression variation depending on the tissue, the physiological state, the developmental stage, and the experimental conditions, as reviewed in [19,24,25]. That is, there are no genes whose expression is stable in all possible materials (cells and tissues) and under any conditions assayed. So, for each experimental design, it will be necessary to test the available genes and, if they show variable expression, to search for new genes and verify that they are suitable as RGs.

In lettuce, novel candidate RGs to study gene expression variation throughout the day and at different developmental stages [26] or in response to different abiotic stresses [27] have been tested. The selection was based on genes that had been previously identified as stable in other species. Nevertheless, thanks to advances in next generation sequencing technologies, it is more and more common to use RNA-seq data as a source of novel candidate RGs, as it has already been done in grapevine, soybean and apple [28–30].

In the present study, stability of six candidate RGs has been evaluated in order to identify suitable genes for the expression analyses of genes related to anthocyanin accumulation in lettuce and wild relatives. Their stability has been analysed in three different experiments: comparison of leaf colour in green and red commercial varieties of lettuce, comparison of tissues (leaf and stem) in a wild relative species, and response to drought stress in cultivated lettuce (commercial and traditional varieties) and in a wild relative species. To the best of our knowledge, this is the first time that the selection of candidate RGs in lettuce has been based on RNA-seq data. The gene expression analyses have not only been carried out in lettuce commercial varieties, but they have also been extended to plant material scarcely studied before [1], e.g., two wild relatives (*Lactuca* spp.) and one lettuce traditional variety.

2. Results

2.1. Selection of Candidate reference genes (RGs) Based on RNA-seq Data

A total of six candidate RGs (Table 1, Table S1) were selected from RNA-seq analyses carried out in *Lactuca* spp. in three different experiments (see "Materials and Methods" section), attending to different selection criteria because classic housekeeping genes previously used in plants did not show stable expression across the different conditions (Table 2). The first one was the expression stability: values of $\log_2$FC (fold change) ranged from −0.75 to 1.03 (Table S2). Then, among those genes that were common in the three experiments, genes with a wide sequence coverage were selected. Finally, one-way analysis of variance (ANOVA) and Student's *t*-test for mean comparison were carried out for gene count data among the groups of each experiment, and six genes

without significant differences (unlike the six classic housekeeping genes analysed) were selected (Table 2). Gene function was also taken into account; five out of six RGs had a structural or metabolic function: cellular component (*ADF2*), ion transport (*CYB5* and *VHA-H*), metabolism (*iPGAM*), and photosynthesis (*TRXL3-3*). However, a probable transcription factor (*SCL13*) was also included, as it is required for hypocotyl elongation regulation during de-etiolation and, therefore, it is not related to anthocyanin content regulation.

Table 1. Description, primer sequences, amplicon length, and qPCR annealing temperature of six candidate reference genes (RGs).

Name	Description	Primer Sequence (5′–3′)	Amplicon Length (bp)	Annealing Temperature (°C)
ADF2	Actin-depolymerizing factor 2	F-TTGGAGAACCAGCAGAAAC R-CCATCAAGCTCTCTCTTGAAC	199	62
CYB5	Cytochrome B5	F-GCACGCTACGAAAGAGG R-CAGGATGATCATCTAGAAAAGG	80	59
iPGAM	Probable 2,3-bisphosphoglycerate -independent phosphoglycerate mutase	F-GGGAGATGTTTCAATTCCAAG R-CCCATTAGAGAAAGATGAGCAG	162	62
SCL13	Scarecrow-like protein 13	F-AGTCGGTTAGCACGGTTA R-TTCGTGTTCGATTCTTGTT	100	56
TRXL3-3	Thioredoxin-like protein 3-3	F-TGGTGTCGTGTTTGTGCAGAG R-GTTGGGTTGTTTCTGGGCATT	112	62
VHA-H	V-type proton ATPase subunit H	F-TGCAAGTGATGATGTTTTGA R-TGCTTGAACAAATGAAGACC	152	59

Table 2. F-values resulting from ANOVA of gene counts for traditional housekeeping genes and new candidate reference genes (RGs) among groups of each experiment: green vs. red, leaf vs. stem, and drought stress.

Gene [a]	Green vs. Red	Leaf vs. Stem	Drought Stress		
			'Romired'	'Morada de Belchite'	*L. homblei*
ACT	0.564 ^{ns}	0.938 ^{ns}	0.421 ^{ns}	0.018 *	0.036 *
α-TUB	0.635 ^{ns}	0.001 **	0.435 ^{ns}	0.637 ^{ns}	0.643 ^{ns}
eEF1-α	0.019 *	0.066 ^{ns}	0.976 ^{ns}	0.089 ^{ns}	0.696 ^{ns}
GAPDH-2C	0.028 *	0.027 *	0.407 ^{ns}	0.302 ^{ns}	0.593 ^{ns}
UBC32	0.159 ^{ns}	0.159 ^{ns}	0.398 ^{ns}	0.276 ^{ns}	0.010 *
UPL6	0.471 ^{ns}	0.013 *	0.003 **	0.013 *	0.168 ^{ns}
ADF2	0.569 ^{ns}	0.884 ^{ns}	0.400 ^{ns}	0.660 ^{ns}	0.547 ^{ns}
CYB5	0.449 ^{ns}	0.270 ^{ns}	0.723 ^{ns}	0.371 ^{ns}	0.673 ^{ns}
iPGAM	0.833 ^{ns}	0.773 ^{ns}	0.128 ^{ns}	0.418 ^{ns}	0.210 ^{ns}
SCL13	0.408 ^{ns}	0.902 ^{ns}	0.883 ^{ns}	0.427 ^{ns}	0.239 ^{ns}
TRXL3-3	0.487 ^{ns}	0.406 ^{ns}	0.843 ^{ns}	0.161 ^{ns}	0.470 ^{ns}
VHA-H	0.198 ^{ns}	0.576 ^{ns}	0.116 ^{ns}	0.195 ^{ns}	0.741 ^{ns}

[a] *ACT*: actin; *α-TUB*: tubulin alpha chain; *eEF1-α*: elongation factor 1-alpha; *GAPDH-2C*: glyceraldehyde-3-phosphate dehydrogenase 2C; *UBC32*: ubiquitin-conjugating enzyme E2 32; *UPL6*: E3 ubiquitin-protein ligase UPL6. ns: non-significant; * $p < 0.05$; ** $p < 0.01$

Some of the RGs that have been traditionally used, such as actin, tubulin, ubiquitin, etc., were planned to be included in this study, but they did not meet all the selection criteria previously described for at least one of the experiments (Table 2).

2.2. Expression Profile of Candidate RGs

Expression levels of the candidate RGs were determined by real-time qPCR using Cq (quantification cycle) data, also referred as Ct (threshold cycle). Cq values of the RGs ranged from 31.38 to 37.82 in green vs. red, from 30.99 to 37.67 in leaf vs. stem, and from 30.43 to 39.17 in the drought stress experiments (Figure 1 and Table S3). *CYB5* and *ADF2* showed the highest expression levels (lowest mean Cq values) in the three experiments, whereas the genes with the lowest expression levels (highest mean Cq values) were *iPGAM*

and *SL13*, except for the wild relative species in the drought stress experiment, that was *TRXL3-3* (followed by those same two).

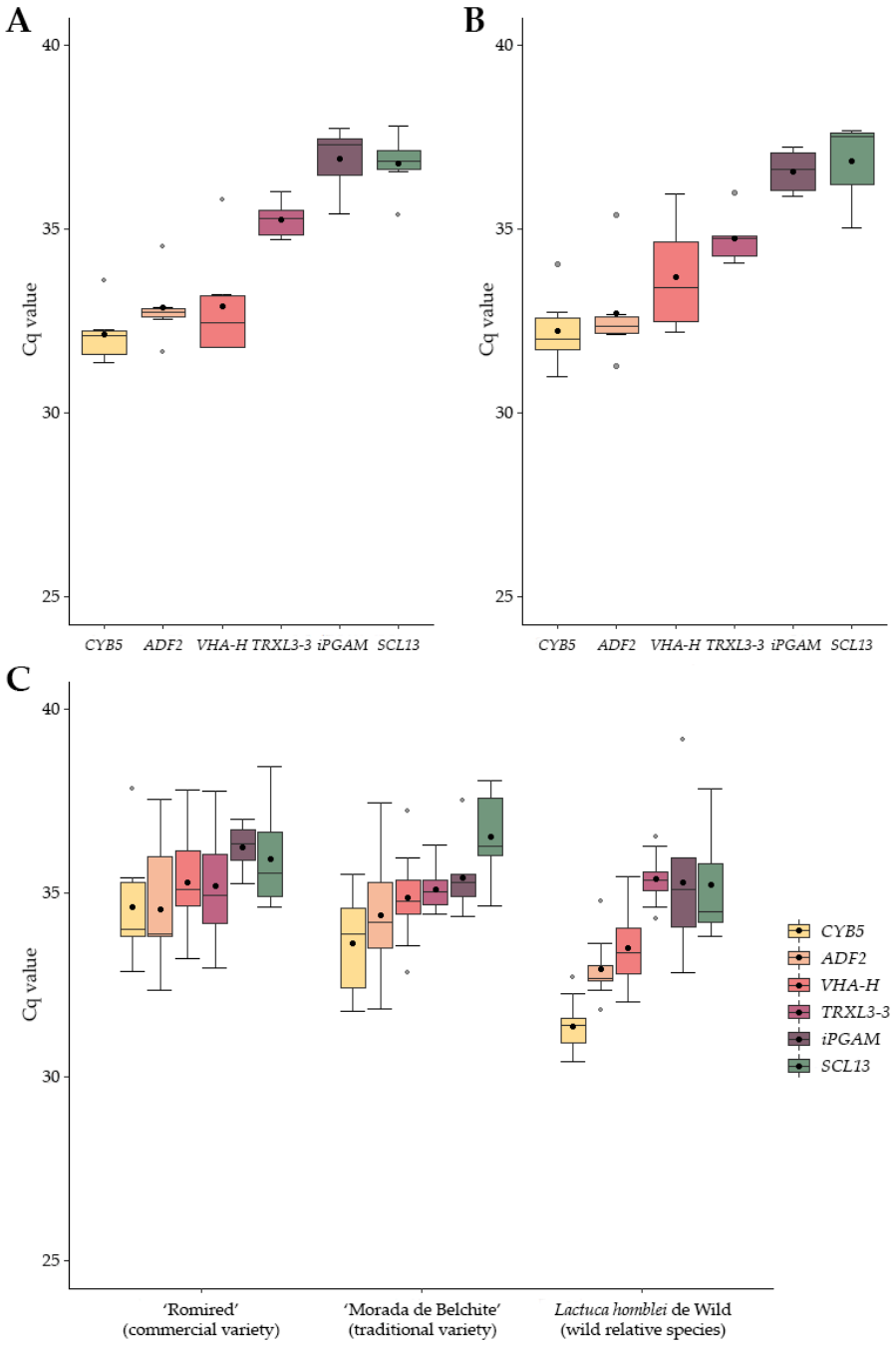


Figure 1. Expression levels, Cq (quantification cycle) values assessed by real-time qPCR (quantitative PCR), of six tested candidate RGs across the samples in three experiments: (**A**) comparison of leaf colour in green and red lettuce commercial varieties; (**B**) comparison of tissues (leaf and stem) in a wild relative species; and (**C**) drought stress in a commercial variety, a traditional variety, and a wild relative species. Lower and upper ends of the boxes represent the 25th and 75th percentiles, respectively, and whisker caps indicate the minimum and maximum values. Horizontal bars and black and grey dots depict the median, mean and outliers, respectively.

The variation width of Cq values across the samples (without considering the outliers) was very different for the six candidate RGs within each experiment (Figure 1). Gene showing the narrowest variation was *ADF2*, except for the drought stress experiment in the cultivated

varieties, in which happened to be *iPGAM* (Figure 1 and Table S3). Nevertheless, a low variation of Cq values is essential but not sufficient to guarantee the stability of RGs expression.

2.3. Analysis of Gene Expression Stability in Accessions of Lactuca: Different Leaf Colour, Tissues, and Drought Stress Conditions

Expression stability of the six candidate RGs in the previously described experiments was assessed by four different analytical tools: geNorm [31], NormFinder [32], and BestKeeper [33] algorithms, and the Delta Ct (ΔCt) method [34], all included in the RefFinder software [35]. M (in geNorm) and SV (stability value, in NormFinder, BestKeeper, and ΔCt) measured the relative expression stability, such that lower M and SV indicated a more stable expression. Nevertheless, each method of analysis follows different statistical procedures (explained in Section 4.5), and the input data are also different, as geNorm and NormFinder algorithms use Cq data corrected with the efficiency values (CqE), whereas BestKeeper and the ΔCt method use mean Cq data. Therefore, stability rankings for each gene within the three experiments show slight differences according to the analytical method applied, as shown in Table 3.

Table 3. Stability values of each candidate reference gene (RG) obtained with geNorm (M), NormFinder, and BestKeeper algorithms, and the Delta Ct method (SV) in three different experiments: green vs. red, leaf vs. stem, and drought stress.

		geNorm		NormFinder		BestKeeper		Delta Ct	
Experiment	**Ranking**	**Gene**	**M**	**Gene**	**SV**	**Gene**	**SV**	**Gene**	**SV**
Green vs. red	1	*CYB5*	0.33	*TRXL3-3*	0.71	*TRXL3-3*	0.40	*CYB5*	0.83
	2	*ADF2*	0.33	*VHA-H*	0.77	*CYB5*	0.53	*ADF2*	0.83
	3	*TRXL3-3*	0.54	*CYB5*	0.81	*SCL13*	0.53	*SCL13*	0.92
	4	*VHA-H*	1.11	*ADF2*	1.09	*ADF2*	0.57	*TRXL3-3*	1.10
	5	*iPGAM*	1.33	*iPGAM*	1.49	*iPGAM*	0.72	*iPGAM*	1.28
	6	*SCL13*	1.90	*SCL13*	2.88	*VHA-H*	1.12	*VHA-H*	1.41
Leaf vs. stem	1	*CYB5*	0.77	*TRXL3-3*	0.30	*TRXL3-3*	0.44	*TRXL3-3*	0.93
	2	*ADF2*	0.77	*VHA-H*	0.30	*iPGAM*	0.47	*CYB5*	1.03
	3	*TRXL3-3*	1.01	*ADF2*	1.18	*CYB5*	0.77	*iPGAM*	1.08
	4	*VHA-H*	1.07	*CYB5*	1.34	*ADF2*	0.89	*ADF2*	1.10
	5	*iPGAM*	1.22	*iPGAM*	1.55	*SCL13*	0.92	*VHA-H*	1.38
	6	*SCL13*	2.41	*SCL13*	4.73	*VHA-H*	1.23	*SCL13*	1.54
Drought stress	1	*TRXL3-3*	1.54	*TRXL3-3*	0.77	*TRXL3-3*	0.85	*SCL13*	3.89
	2	*ADF2*	1.54	*ADF2*	0.77	*SCL13*	0.92	*TRXL3-3*	3.90
	3	*CYB5*	1.98	*iPGAM*	2.81	*iPGAM*	1.11	*iPGAM*	4.01
	4	*iPGAM*	3.40	*CYB5*	2.91	*ADF2*	1.23	*ADF2*	4.04
	5	*SCL13*	4.41	*SCL13*	6.62	*CYB5*	1.50	*CYB5*	4.07
	6	*VHA-H*	6.82	*VHA-H*	11.27	*VHA-H*	10.35	*VHA-H*	14.08

For the comparison of gene expression in green vs. red leaf colour (Figure 2A and Table 3), the two most stable genes were *CYB5* and *ADF2*, *TRXL3-3* and *VHA-H*, *TRXL3-3* and *CYB5*, and *CYB5* and *ADF2* according to geNorm, NormFinder, BestKeeper and ΔCt, respectively. By contrast, the gene with the lowest stability value was *SCL13* according to geNorm and NormFinder, and *VHA-H* according to BestKeeper and ΔCt, being *iPGAM* found the second least stable gene with the four analytical tools.

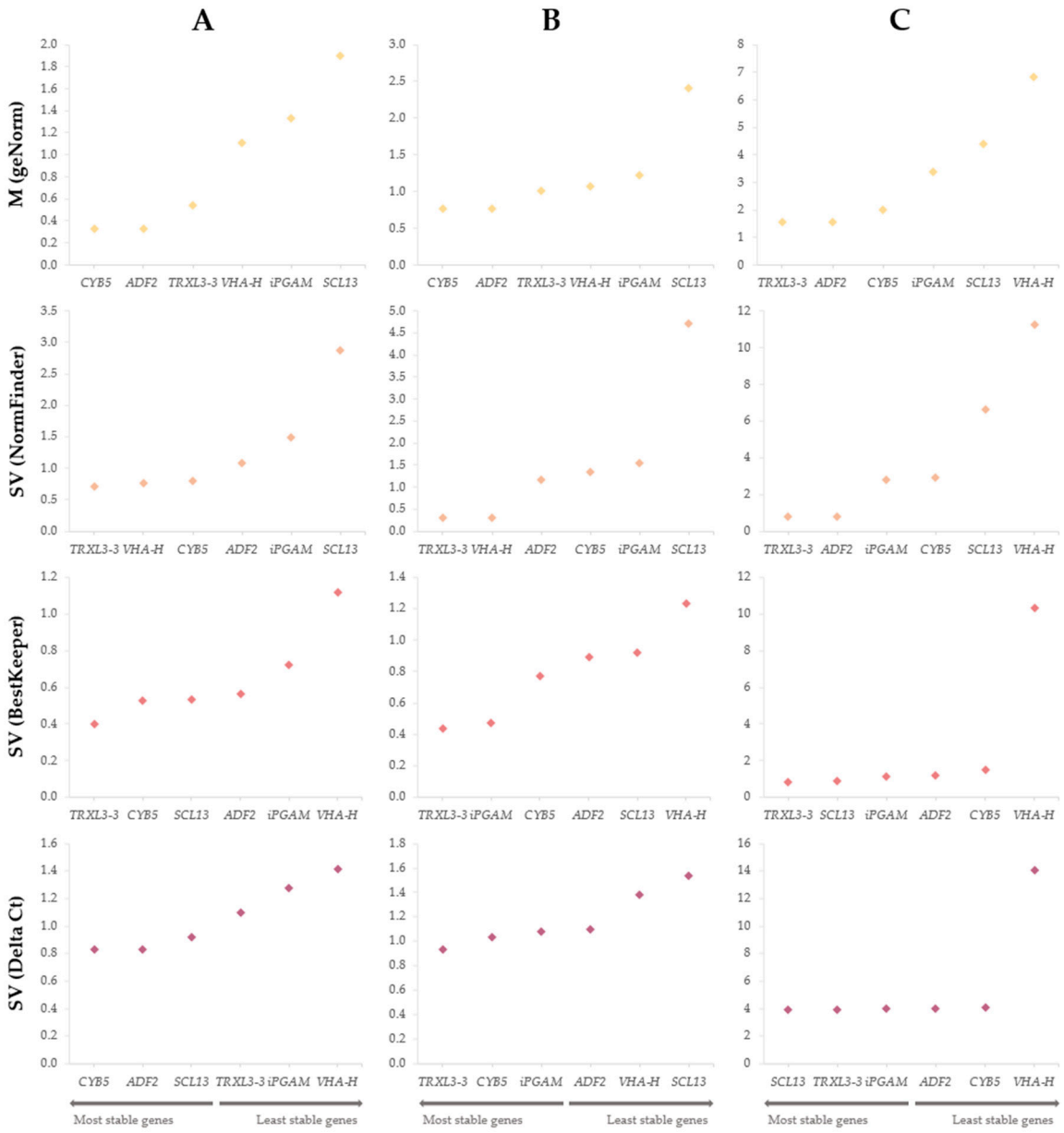


Figure 2. Expression stability of six tested candidate reference genes (RGs) calculated by (♦) geNorm (M), (♦) NormFinder, (♦) BestKeeper, and (♦) ΔCt (SV) methods in three experiments: (**A**) comparison of leaf colour (green and red) in lettuce commercial varieties; (**B**) comparison of tissues (leaf and stem) in a wild relative species; and (**C**) under drought stress in a commercial variety, a traditional variety, and a wild relative. The most stable genes are represented on the left and the least stable on the right of the graph.

When comparing leaf vs. stem tissues (Figure 2B and Table 3), the top two positions in the ranking were occupied by *CYB5* and *ADF2*, *TRXL3-3* and *VHA-H*, *TRXL3-3* and *iPGAM*, and *TRXL3-3* and *CYB5*, applying geNorm, NormFinder, BestKeeper and the ΔCt method, respectively. In contrast, *SCL13* was ranked at the bottom according to geNorm, NormFinder and ΔCt, whereas *VHA-H* was the most unstable gene according to BestKeeper.

Finally, in the drought stress experiment (Figure 2C and Table 3), the most stable genes resulted to be *TRXL3-3* and *ADF2*, according to geNorm and NormFinder, and *TRXL3-3* and *SCL13* (though in reverse order), according to BestKeeper and ΔCt. In this case, *VHA-H* showed the lowest expression stability according to the four methods of analysis.

Due to the slight differences observed when using the different analytical tools, a comprehensive ranking that integrates stability values calculated with all of them was created for each experiment (Table 4), following the procedure described in Section 4.5. Thus, the most stable RG for green vs. red was *CYB5* followed by *TRXL3-3*. These two genes reached the two top positions again in the leaf vs. stem experiment, although in the opposite order. Finally, *TRXL3-3* was also found to be the best RG for the drought stress experiment followed, in this case, by *ADF2*. On the other hand, *iPGAM*, *SCL13*, and *VHA-H* resulted to be the most unstable genes for green vs. red, leaf vs. stem, and drought stress experiments, respectively.

Table 4. Comprehensive stability ranking of six candidate reference genes (RGs) for three different experiments: green vs. red, leaf vs. stem, and drought stress.

Ranking	Green vs. Red	Leaf vs. Stem	Drought Stress
1	*CYB5*	*TRXL3-3*	*TRXL3-3*
2	*TRXL3-3*	*CYB5*	*ADF2*
3	*ADF2*	*ADF2*	*SCL13*
4	*VHA-H*	*VHA-H*	*iPGAM*
5	*SCL13*	*iPGAM*	*CYB5*
6	*iPGAM*	*SCL13*	*VHA-H*

Two more comprehensive rankings were created considering separately analytical methods that use either CqE (geNorm and NormFinder) or Cq (BestKeeper and ΔCt) data (Table S4). In comparison with the ranking that comprises results from the four analytical tools (Table 4), the genes at the highest positions coincided, except in the case of the BestKeeper + ΔCt ranking for the drought stress experiment. Similarly, the genes in the lowest positions were mostly the same in the case of geNorm + Normfinder ranking (in reverse order in the leaf colour experiment) whereas they were slightly different for the BestKeeper + ΔCt ranking in two out of the three experiments (leaf colour and tissues).

3. Discussion

Robust normalisation of real-time qPCR data by RGs is essential to obtain reliable results in gene expression studies. Furthermore, it is well known that identification of stable RGs is necessary for each experimental context. In this work, the suitability of six candidate RGs has been evaluated in expression studies of anthocyanin-related genes in three different experiments: comparison of leaf colour (green and red) in lettuce commercial varieties, comparison of tissues (leaf and stem) in a wild relative, and drought stress in two lettuce varieties (one commercial and one traditional) and a wild relative.

Expression of genes with an effect on anthocyanin content has been scarcely studied in lettuce, except in red-leaf varieties in normal conditions [3–7]. The first step in this kind of analyses should be the selection of suitable RGs. Identification of stable genes as candidate RGs in lettuce has been previously tackled to assess gene expression throughout the day and in different developmental stages [26], and in the response to abiotic stresses [27]. Although RGs for the normalization of expression data under drought stress have already been identified, the target gene tested to validate the selected RGs was not directly related to anthocyanin accumulation [27]. In addition, RGs for the other two experiments (green vs. red and leaf vs. stem) and the plant material studied here have not been tested before, as far as we know. Furthermore, unlike in these previous studies in lettuce, selection of candidate RGs in this work was based on RNA-seq data. Through the bioinformatic analysis of these RNA-seq data, we selected a set of candidate RGs (*ADF2*, *CYB5*, *iPGAM*, *TRXL3-3*, *SCL13*, and *VHA-H*) that were different from those tested in [26,27]. Classic housekeeping genes used in plants [20–23] were discarded since in general they did not meet all our selection criteria (i.e., they were not stably expressed or the sequence coverage was insufficient or ANOVA F-value < 0.05) in at least one of the experiments. Therefore, it is the first time that this set of candidate RGs is tested in lettuce, to the best of our knowledge.

Four different analytical tools have been used to assess the six candidate RG stability and then a comprehensive ranking was elaborated to integrate the results from the four methods. According to this ranking, the two most stable genes, that is, the most suitable as RGs, were *CYB5* and *TRXL3-3* for green vs. red experiment, again *CYB5* and *TRXL3-3* for the leaf vs. stem experiment, and *TRXL3-3* followed by *ADF2* for the drought stress experiment (Table 4). In contrast, the most unstable genes and therefore, the less suited to be used as RGs, were *iPGAM*, *SCL13,* and *VHA-H*, for each of the three experiments above mentioned, respectively (Table 4).

Most of the tested candidate RGs have a structural or metabolic function, except for *SCL13*, a probable transcription factor (Table 1). Even so, *SCL13* was selected because firstly, it revealed itself stable according to the RNA-seq results, and secondly, it participates in the regulation of the hypocotyl elongation during the seedling de-etiolation, that is a totally different process from those of our interest (anthocyanin biosynthesis/degradation and transport). However, in general, it exhibited a low position in the comprehensive stability ranking in our experiments (Table 4). In spite of being a regulatory element, its low stability might be due to the qPCR efficiency values obtained (Table S3) and not to its function as a transcription factor. This is supported by the fact that *SCL13* positions were generally higher in the rankings that did not take into account efficiency data (they used directly mean Cq values) and, hence, they were lower when efficiency data were considered (rankings that used CqE values) (Table 3). This happened for the three experiments, but the differences especially stood out in the drought stress experiment, where *SCL13* position dropped from the first in Bestkeeper + ΔCt ranking to the third in geNorm + NormFinder and the global rankings (Table 4 and Table S4). That is, correction with efficiency values makes *SCL13* become a less stable gene.

It is known that results of gene expression assays are hugely influenced by qPCR efficiency, and in turn, qPCR efficiency is very dependent on the primer design [36]. In our study, some efficiency values were relatively low (Table S3) as there were important sequence limitations during the primer design (Table S1). The two wild relative species included in this study, *Lactuca homblei* de Wild and *Lactuca squarrosa* (Thunb.) Miq., belong to the tertiary lettuce gene pool [37], so they are genetically distant from the cultivated lettuce, *L. sativa*. In the case of the drought stress experiment, the use of the same pairs of primers for *L. sativa* varieties and *L. homblei* was mandatory, whereas in the tissue experiment, different primers could have been designed though we opted for using the same to be able to establish comparisons among all the experiments, assuming some penalization in terms of efficiency. Therefore, a redesign of primers in a study without our limitations would probably improve the efficiency values and increase the expression stability of the candidate RGs that were negatively affected by it.

The rankings obtained using CqE data (geNorm + NormFinder) are more similar to the global one (Table 4) than those generated with Cq values (BestKeeper + ΔCt) (Table S4). This supports the importance of considering efficiency data, as it has already been reported in other crops like wheat [38]. So, we recommend the use of geNorm and NormFinder algorithms over BestKeeper and the ΔCt method. Indeed, in the last few years, geNorm and NormFinder seems to be the most frequently used tools to measure the RG stability in gene expression studies, especially in detriment of the ΔCt method. Nevertheless, in our study, differences among the three rankings are not important since those RGs identified as the most or least stable occupied generally the upper or lower positions, respectively (Tables 4 and S4).

4. Materials and Methods

4.1. Plant Material and Experimental Designs

Three different experiments were performed using especially appropriate plant materials in each case (Table 5). First, to compare leaf colour, the commercial varieties 'Begoña' and 'Romired', green and red lettuces, respectively, were cultivated in winter 2018/2019. Second, to study different tissues (leaf and stem), a wild relative species with bushy growth habit, *L.*

squarrosa, was grown in winter 2020/2021. Third, a drought stress experiment was carried out with the commercial variety 'Romired', the traditional variety 'Morada de Belchite', and the wild relative species *L. homblei*, cultivated in winter 2020/2021. Concretely, this experiment consisted of three different irrigation schedulings applied three weeks before harvesting: C (control or full irrigation, week 1: 1350 mL, weeks 2–3: 2100 mL/each), DI-1 (Deficit Irrigation 1, week 1: 450 mL, weeks 2–3: 150 mL/each), and DI-2 (weeks 1–3: 0 mL). In the leaf colour experiment, 3 plants per accession were cultivated in a greenhouse at Ramiro Arnedo S.A. (Almería, Spain). In the tissue experiment, 3 plants per accession were cultivated in a greenhouse at Agrifood Research and Technology Centre of Aragón (CITA, Zaragoza, Spain). In the drought stress experiment, 3 plants per accession for each of the 3 treatments (9 plants per accession in total) were cultivated in a greenhouse at CITA. In all cases, the plants were distributed following a complete randomized block design. Plants were grown in pots (30 × 25 cm and 11.7 L volume) with a mix of black and blonde peat (1:1) with fertilizer incorporated but without supplemental lighting to avoid the stimulation of anthocyanin synthesis. After a period ranging from 3 (2018/2019) to 3 and a half months (2020/2021), two leaves per plant (one inner and one outer) were collected, as well as parts of the stem in the case of *L. squarrosa*, so that the whole plant was represented. The samples were immediately frozen with liquid nitrogen and then kept at −80 °C until their lyophilization.

Table 5. Plant material used in the three experiments of the present study.

Experiment	Accession Name	Species	Group	Leaf Colour	Year	Source [a]	Accession Number
Leaf colour (green vs. red)	'Begoña'	*Lactuca sativa* L.	Commercial variety	Green	2018/2019	Ramiro Arnedo Semillas S.A.	-
	'Romired'	*Lactuca sativa* L.	Commercial variety	Red		CGN	CGN24713
Tissue (leaf vs. stem)	*Lactuca squarrosa*	*Lactuca squarrosa* (Thunb.) Miq.	Wild crop relative	Semi-red (red stems)	2020/2021	BGHZ	BGHZ5124
Drought stress (C vs. DI-1 vs. DI-2) [b]	'Romired'	*Lactuca sativa* L.	Commercial variety	Red	2020/2021	CGN	CGN24713
	'Morada de Belchite'	*Lactuca sativa* L.	Traditional variety	Semi-red		BGHZ	BGHZ0527
	Lactuca homblei	*Lactuca homblei* De Wild	Wild crop relative	Semi-red		BGHZ	BGHZ5322

[a] BGHZ: Vegetable Germplasm Bank of Zaragoza (Spain); CGN: Centre for Genetic Resources (Wageningen, The Netherlands). [b] C: Control (full irrigation, week 1: 1350 mL, weeks 2–3: 2100 mL/each), DI-1 (Deficit Irrigation 1, week 1: 450 mL, weeks 2–3: 150 mL/each); DI-2 (Deficit Irrigation 2, weeks 1–3: 0 mL).

4.2. RNA Extraction and RNA-Seq Analysis

Total RNA was extracted from lyophilized samples using the NZY Total RNA Isolation kit (NZYtech Lda.-Genes and Enzymes, Lisbon, Portugal) and then treated with DNase to ensure elimination of DNA using the TURBO DNA-free™ kit (Invitrogen, Waltham, MA, USA), according to manufacturers' instructions. Quantity and purity of extracted RNA were measured with the Eukaryotic Total RNA Nanobioanalyzer Assay in a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The obtained RNA samples were used to build a total of 39 cDNA libraries resulting from the 3 experiments (leaf colour: 2 accessions × 3 replicates; plant tissue: 2 tissues × 3 replicates; drought stress: 3 accessions × 3 replicates × 3 irrigation regimes). All libraries were sequenced in both directions using the TruSeq Stranded mRNA protocol (Illumina, San Diego, CA, USA) with a NovaSeq 6000 S1 instrument (Illumina) to obtain from 32 to 111 million strand-specific pair-end reads of 100 base pairs (bp) each. Sequencing was performed at the National Centre for Genomic Regulation (CNAG-CRG, Barcelona, Spain).

Sequence analyses were performed using the Galaxy tool [39]. Adapter sequences were removed using Trimmomatic (Galaxy Version 0.38.1) [40] and reads were aligned to the lettuce reference genome Last_Salinas_V7 (GCF_002870075.2) using HISAT2 (Galaxy Version 2.2.1+galaxy0) [41], with a maximum length of 20,000 bp between exons. MarkDuplicates (Galaxy Version 2.18.2.2) and FixMateInformation (Galaxy Version 2.18.2.1) Picard tools [42] were used to filter out optical duplicates and to confirm mate-pairs, respectively. Read

counts were generated using featureCounts (Galaxy Version 2.0.1+galaxy2) [43] and differential expression analyses were performed using edgeR (Galaxy Version 3.36.0+galaxy0) [44].

4.3. Selection of Candidate RGs

RNA-seq data were used to select the candidate RGs among those that met the criteria explained bellow. Firstly, all genes must show stable expression across accessions, tissues, or treatments within each assay. So, they were filtered for low values of the FC (fold change, $\log_2(FC) \leq 1$) as values close to zero indicate that there is neither an increase nor a decrease in the gene expression among the groups compared, and the FDR (False Discovery Rate) threshold of 5% (adjusted *p*-value via the Benjamini and Hochberg method using the edgeR package [44]). Secondly, genes that presented a wide sequence coverage were selected. Finally, one-way analysis of variance (ANOVA) and Student's *t*-test for mean comparison were performed for the gene counts among the groups within each experiment with JMP v5.1.2 software for Windows (SAS Institute Inc., Cary, NC, USA). Eventually, taking also into account their biological functions, six genes without significant differences were chosen: *ADF2*, *CYB5*, *iPGAM*, *SCL13*, *TRXL3-3*, and *VHA-H* (Table 1). They were mainly genes involved in constitutive processes, as they are expected to have stable expression in any tissue and under any circumstances. Specifically, they are genes encoding for cytoskeleton components (*ADF2*), enzymes catalysing metabolic reactions (*iPGAM* and *TRXL3-3*), and electron (*CYB5*) or proton (*VHA-H*) transporters. However, a probable transcriptional regulator of seed germination (*SCL13*) was also included with the intention of testing whether regulatory elements could be used as RGs provided that they do not participate in the same processes under study.

4.4. mRNA Isolation, cDNA Synthesis and Real-Time qPCR

mRNA purification from total RNA samples described above was performed using the Dynabeads mRNA DIRECT™ kit (Invitrogen), followed by the cDNA synthesis using the NZY M-MuLV First-Strand cDNA Synthesis separate oligos kit (NZYTech), as recommended by manufacturers. Specific pairs of primers were designed for each selected candidate RG (Table 1) using OLIGO software version 6.45 (Cascade, CO, USA) and a consensus sequence from the three species under study as template, the cultivated lettuce (*L. sativa*), and the two wild relatives (*L. homblei* and *L. squarrosa*) were selected, paying special attention to exclude any ambiguity in their sequences. Real-time qPCR reactions were performed on a StepOnePlus™ System (Applied Biosystems, Waltham, MA, USA). Each reaction was run in a final volume of 12 µL containing 1 µL of 1:40 diluted cDNA, 0.40 µM of forward and reverse primers (Integrated DNA Technologies, IDT, Coralville, Iowa, USA), and 1x NZYSupreme qPCR Green Master Mix, ROX plus (NZYTech). The amplification conditions were the following: 2 min at 95 °C and 40 cycles of 5 s at 95 °C, 15 s at 56–62 °C (Table 1) and 30 s at 72 °C. Melting curve analyses were carried out to verify the specificity of each reaction and it ranged from 72 °C to 95 °C with 0.3 °C increment per cycle. Two technical replicates per sample were performed and non-template controls were added to ensure that contamination with genomic DNA had not occurred.

4.5. Stability Analysis of RGs

To evaluate gene expression stability of the set of RGs, the data were analysed with geNorm [31], NormFinder [32], and BestKeeper [33] algorithms, and the comparative ΔCt method [34], all of them integrated in RefFinder software [35]. geNorm algorithm calculates the expression stability value (M value) based on the average pairwise variation ($V_{n/n+1}$) for a candidate RG with all the other tested genes. Lower M values indicate more stable expression. NormFinder uses an ANOVA-based model that calculates SV considering intra- and inter-group variation. Lower values represent more stability. BestKeeper ranks the stability of candidate RGs considering three variables: standard deviation (SD), coefficient of variation (CV) and correlation coefficient (r). The lower values of SD and CV and the higher values of r, the more stable is the gene expression (lower SV). The comparative ΔCt method first obtains

the difference between the Cq of the treated and control samples. Then, it displays pairwise comparisons between the genes by calculating the mean SD. Lower values of SD, and hence of SV, indicate higher stability of the genes. For BestKeeper and ΔCt, mean values of Cq were used, whereas for geNorm and NormFinder, corrected values of Cq with the efficiency data were used, according to the formula CqE = Cq*(log(E)/log(2)) [45]. In all cases, mean data of two technical replicates per sample were used.

Finally, a comprehensive ranking that comprises results from the four analytical tools (geNorm, NormFinder, BestKeeper, and the comparative ΔCt method) was elaborated for each of the three experiments. Firstly, individual rankings were created by assigning a certain weight to each gene according to the results obtained from each of these statistical methods. Then, a final weighted list was obtained by calculating geometric means from the four individual rankings for each RG. In addition, two more comprehensive rankings for each experiment were obtained following the same procedure, one comprising stability data from geNorm and NormFinder, and the other from BestKeeper and ΔCt, methods that use corrected and non-corrected values of Cq, respectively.

5. Conclusions

In this study, a novel set of six candidate RGs with low variation in expression levels in *Lactuca* spp. were selected based on an RNA-seq data analysis. Their RNA-seq expression was validated by qPCR and candidate RGs were subjected to analyses of expression stability through different analytical tools. We have found stable genes, hence suitable RGs for an accurate real-time qPCR normalization in three different experiments in the genus *Lactuca*. For studying the expression of anthocyanin-related genes in these *Lactuca* spp., we concretely recommend *CYB5* for the experiment comparing the green and red-leaf commercial varieties of lettuce; and *TRXL3-3* for the comparison of tissues (leaf and stem) in a wild relative species with bushy growth, and for the drought stress experiment in two cultivated varieties of lettuce (one commercial and one traditional) and in a wild relative. On the other hand, *iPGAM*, *SCL13*, and *VHA-H* should be avoided in all three experiments set out here, although an increase in efficiency values by primer redesign (when possible) might improve expression stability in some cases. Furthermore, the six candidate RGs studied here, but especially the two revealed as the most stable (*CYB5* and *TRXL3-3*), could be tested to verify if they show stable expression in any *Lactuca* spp. as long as the processes under study are not related to those aforementioned in which these six genes participate. The results of this study provide valuable information for future gene expression studies in different accessions of lettuce, different tissues, and conditions of stress.

In addition, these stable RGs will be used as internal controls in future research on expression of genes related to anthocyanin content, among which we expect to find biosynthesis genes, genes encoding proteins of transport and transcription factors within each experiment: comparison of leaf colour, of tissues, and drought stress, respectively.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms24033052/s1.

Author Contributions: Conceptualization, A.D.; methodology, A.D., J.G. and I.M.-L.; software, J.G., and I.M.-L.; validation, A.D., J.G. and I.M.-L.; formal analysis, A.D., J.G. and I.M.-L.; investigation, A.D. and I.M.-L.; resources, A.D., I.M.-L. and M.S.A.; data curation, J.G. and I.M.-L.; writing—original draft preparation, A.D. and I.M.-L.; writing—review and editing, A.D., I.M.-L., J.G. and M.S.A.; visualization, I.M.-L.; supervision, A.D. and J.G.; project administration, A.D.; funding acquisition, A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the projects RTA2017-00093-00-00 from the National Institute for Agricultural and Food Research and Technology (INIA), LMP164_18 and LMP148_21 from the Government of Aragón, by the Operational Programmes FEDER Aragón 2014-2020 and 2020-2022, and by the European Social Fund from the European Union (A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética"). I.M.-L. was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities

(MCIU) and the Spanish State Research Agency (AEI). The APC was funded by the aforementioned projects and the University of Zaragoza (Unizar) through the AgriFood Institute of Aragon (IA2).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data generated during the study are within the article or supplementary material.

Acknowledgments: We thank J. A. Aranjuelo for technical support. We gratefully acknowledge the Vegetable Germplasm Bank of Zaragoza (BGHZ-CITA, Spain) and the Centre for Genetic Resources (CNG, Wageningen, The Netherlands) for supplying the seeds used for this work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Medina-Lozano, I.; Bertolín, J.R.; Díaz, A. Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin C and anthocyanin content. *Food Chem.* **2021**, *359*, 129864. [CrossRef]
2. Yousuf, B.; Gul, K.; Wani, A.A.; Singh, P. Health Benefits of Anthocyanins and Their Encapsulation for Potential Use in Food Systems: A Review. *Crit. Rev. Food Sci. Nutr.* **2016**, *56*, 2223–2230. [CrossRef] [PubMed]
3. Moreno-Escamilla, J.O.; Jimeńez-Hernández, F.E.; Alvarez-Parrilla, E.; De La Rosa, L.A.; MartÍnez-Ruiz, N.D.R.; González-Fernández, R.; Orozco-Lucero, E.; González-Aguilar, G.A.; García-Fajardo, J.A.; Rodrigo-García, J. Effect of Elicitation on Polyphenol and Carotenoid Metabolism in Butterhead Lettuce (*Lactuca sativa* var. capitata). *ACS Omega* **2020**, *5*, 11535–11546. [CrossRef] [PubMed]
4. Su, W.; Tao, R.; Liu, W.; Yu, C.; Yue, Z.; He, S.; Lavelle, D.; Zhang, W.; Zhang, L.; An, G.; et al. Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* **2020**, *18*, 479–490. [CrossRef] [PubMed]
5. Wada, K.C.; Inagaki, N.; Sakai, H.; Yamashita, H.; Nakai, Y.; Fujimoto, Z.; Yonemaru, J.; Itoh, H. Genetic effects of *Red Lettuce Leaf* genes on red coloration in leaf lettuce under artificial lighting conditions. *Plant-Environ. Interact.* **2022**, *3*, 179–192. [CrossRef]
6. Zhang, Y.Z.; Xu, S.Z.; Cheng, Y.W.; Ya, H.Y.; Han, J.M. Transcriptome analysis and anthocyanin-related genes in red leaf lettuce. *Genet. Mol. Res.* **2016**, *15*, 10-4238. [CrossRef] [PubMed]
7. Zhang, L.; Su, W.; Tao, R.; Zhang, W.; Chen, J.; Wu, P.; Yan, C.; Jia, Y.; Larkin, R.M.; Lavelle, D.; et al. RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat. Commun.* **2017**, *8*, 2264. [CrossRef]
8. Ren, W.; Zhao, L.; Zhang, L.; Wang, Y.; Cui, L.; Tang, Y.; Sun, X.; Tang, K. Molecular cloning and characterization of 4-hydroxyphenylpyruvate dioxygenase gene from *Lactuca sativa*. *J. Plant Physiol.* **2011**, *168*, 1076–1083. [CrossRef]
9. Damerum, A.; Selmes, S.L.; Biggi, G.F.; Clarkson, G.J.J.; Rothwell, S.D.; Truco, M.J.; Michelmore, R.W.; Hancock, R.D.; Shellcock, C.; Chapman, M.A.; et al. Elucidating the genetic basis of antioxidant status in lettuce (*Lactuca sativa*). *Hortic. Res.* **2015**, *2*, 15055. [CrossRef]
10. Park, S.; Shi, A.; Mou, B. Genome-wide identification and expression analysis of the *CBF/DREB1* gene family in lettuce. *Sci. Rep.* **2020**, *10*, 5733. [CrossRef]
11. Chen, L.; Xu, M.; Liu, C.; Hao, J.; Fan, S.; Han, Y. *LsMYB15* Regulates Bolting in Leaf Lettuce (*Lactuca sativa* L.) Under High-Temperature Stress. *Front. Plant Sci.* **2022**, *13*, 921021. [CrossRef]
12. Liu, R.; Su, Z.; Zhou, H.; Huang, Q.; Fan, S.; Liu, C.; Han, Y. LsHSP70 is induced by high temperature to interact with calmodulin, leading to higher bolting resistance in lettuce. *Sci. Rep.* **2020**, *10*, 15155. [CrossRef]
13. Xiong, T.; Zhang, S.; Kang, Z.; Zhang, T.; Li, S. Dose-Dependent Physiological and Transcriptomic Responses of Lettuce (*Lactuca sativa* L.) to Copper Oxide Nanoparticles-Insights into the Phytotoxicity Mechanisms. *Int. J. Mol. Sci.* **2021**, *22*, 3688. [CrossRef] [PubMed]
14. Zhao, J.; Lu, C.; Tariq, M.; Xiao, Q.; Zhang, W.; Huang, K.; Lu, Q.; Lin, K.; Liu, Z. The response and tolerance mechanisms of lettuce (*Lactuca sativa* L.) exposed to nickel in a spiked soil system. *Chemosphere* **2019**, *222*, 399–406. [CrossRef] [PubMed]
15. Porcel, R.; Aroca, R.; Azcón, R.; Ruiz-Lozano, J.M. PIP Aquaporin Gene Expression in Arbuscular Mycorrhizal *Glycine max* and *Lactuca sativa* Plants in Relation to Drought Stress Tolerance. *Plant Mol. Biol.* **2006**, *60*, 389–404. [CrossRef]
16. Ruiz-Lozano, J.M.; Aroca, R.; Zamarreño, Á.M.; Molina, S.; Andreo-Jiménez, B.; Porcel, R.; García-Mina, J.M.; Ruyter-Spira, C.; López-Ráez, J.A. Arbuscular mycorrhizal symbiosis induces strigolactone biosynthesis under drought and improves drought tolerance in lettuce and tomato. *Plant Cell Environ.* **2016**, *39*, 441–452. [CrossRef]
17. Aksakal, O.; Tabay, D.; Esringu, A.; Icoglu Aksakal, F.; Esim, N. Effect of proline on biochemical and molecular mechanisms in lettuce (*Lactuca sativa* L.) exposed to UV-B radiation. *Photochem. Photobiol. Sci.* **2017**, *16*, 246–254. [CrossRef] [PubMed]
18. Navarro, E.; Serrano-Heras, G.; Castaño, M.J.; Solera, J. Real-time PCR detection chemistry. *Clin. Chim. Acta* **2015**, *439*, 231–250. [CrossRef] [PubMed]
19. Kozera, B.; Rapacz, M. Reference genes in real-time PCR. *J. Appl. Genet.* **2013**, *54*, 391–406. [CrossRef]

20. Jain, M.; Nijhawan, A.; Tyagi, A.K.; Khurana, J.P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 646–651. [CrossRef]
21. Kiarash, J.G.; Dayton Wilde, H.; Amirmahani, F.; Mehdi Moemeni, M.; Zaboli, M.; Nazari, M.; Saeed Moosavi, S.; Jamalvandi, M. Selection and validation of reference genes for normalization of qRT-PCR gene expression in wheat (*Triticum durum* L.) under drought and salt stresses. *J. Genet.* **2018**, *97*, 1433–1444. [CrossRef] [PubMed]
22. Liu, D.; Shi, L.; Han, C.; Yu, J.; Li, D.; Zhang, Y. Validation of Reference Genes for Gene Expression Studies in Virus-Infected *Nicotiana benthamiana* Using Quantitative Real-Time PCR. *PLoS ONE* **2012**, *7*, e46451. [CrossRef] [PubMed]
23. de Jesus Miranda, V.; Coelho, R.R.; Viana, A.A.B.; de Oliveira Neto, O.B.; Carneiro, R.M.D.G.; Rocha, T.L.; de Sa, M.F.G.; Fragoso, R.R. Validation of reference genes aiming accurate normalization of qPCR data in soybean upon nematode parasitism and insect attack. *BMC Res. Notes* **2013**, *6*, 196. [CrossRef]
24. Gutierrez, L.; Mauriat, M.; Guénin, S.; Pelloux, J.; Lefebvre, J.-F.; Louvet, R.; Rusterucci, C.; Moritz, T.; Guerineau, F.; Bellini, C.; et al. The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J.* **2008**, *6*, 609–618. [CrossRef] [PubMed]
25. Joseph, J.T.; Poolakkalody, N.J.; Shah, J.M. Plant reference genes for development and stress response studies. *J. Biosci.* **2018**, *43*, 173–187. [CrossRef]
26. Sgamma, T.; Pape, J.; Massiah, A.; Jackson, S. Selection of reference genes for diurnal and developmental time-course real-time PCR expression analyses in lettuce. *Plant Methods* **2016**, *12*, 21. [CrossRef] [PubMed]
27. Borowski, J.M.; Galli, V.; da Silva Messias, R.; Perin, E.C.; Buss, J.H.; dos Anjos e Silva, S.D.; Rombaldi, C.V. Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. *Planta* **2014**, *239*, 1187–1200. [CrossRef]
28. Song, Y.; Hanner, R.H.; Meng, B. Genome-wide screening of novel RT-qPCR reference genes for study of GLRaV-3 infection in wine grapes and refinement of an RNA isolation protocol for grape berries. *Plant Methods* **2021**, *17*, 110. [CrossRef]
29. Yim, A.K.Y.; Wong, J.W.H.; Ku, Y.S.; Qin, H.; Chan, T.F.; Lam, H.M. Using RNA-seq Data to Evaluate Reference Genes Suitable for Gene Expression Studies in Soybean. *PLoS ONE* **2015**, *10*, e0136343. [CrossRef]
30. Zhou, Z.; Cong, P.; Tian, Y.; Zhu, Y. Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. *PLoS ONE* **2017**, *12*, e0185288. [CrossRef]
31. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, *3*, research0034.1. [CrossRef] [PubMed]
32. Andersen, C.L.; Jensen, J.L.; Ørntoft, T.F. Normalization of Real-Rime Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. *Cancer Res.* **2004**, *64*, 5245–5250. [CrossRef] [PubMed]
33. Pfaffl, M.W.; Tichopad, A.; Prgomet, C.; Neuvians, T. Determination of most stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **2004**, *26*, 509–515. [CrossRef]
34. Silver, N.; Best, S.; Jiang, J.; Thein, S.L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* **2006**, *7*, 33. [CrossRef]
35. Xie, F.; Xiao, P.; Chen, D.; Xu, L.; Zhang, B. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **2012**, *80*, 75–84. [CrossRef] [PubMed]
36. Bustin, S.A.; Benes, V.; Garson, J.A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.W.; Shipley, G.L.; et al. The MIQE Guidelines: *M*inimum *I*nformation for Publication of *Q*uantitative Real-Time PCR *E*xperiments. *Clin. Chem.* **2009**, *55*, 611–622. [CrossRef]
37. PGR (Plant Genetic Resources) Lettuce. The Lettuce Gene Pool. Available online: https://www.pgrportal.nl/en/lettuce-genetic-resources-portal.htm (accessed on 21 November 2022).
38. Garrido, J.; Aguilar, M.; Prieto, P. Identification and validation of reference genes for RT-qPCR normalization in wheat meiosis. *Sci. Rep.* **2020**, *10*, 2726. [CrossRef]
39. Afgan, E.; Baker, D.; Batut, B.; Van Den Beek, M.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Grüning, B.A.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [CrossRef]
40. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
41. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
42. Broad Institute. Picard Tools. Available online: http://broadinstitute.github.io/picard/ (accessed on 25 October 2022).
43. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2013**, *30*, 923–930. [CrossRef] [PubMed]

44. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140. [CrossRef] [PubMed]
45. Kubista, M.; Sindelka, R. The Prime Technique-Real-time PCR Data Analysis. *G.I.T Lab. J.* **2007**, *9*, 33–35.

CHAPTER 4.1. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.mdpi.com/article/10.3390/ijms24033052/s1.

Table S1. Characteristics of the primer pairs used in the real-time qPCR (quantitative PCR) reactions.

Name	Primer sequence (5' - 3')	F 3'-Duplex formation (ΔG)	R 3'- Duplex formation (ΔG)	F-R 3'-Duplex formation (ΔG)	F Hairpin (maximum ΔG)‡	R Hairpin (maximum ΔG) ‡	F Internal stability (5' - 3')	R Internal stability (5' - 3')	Tm (°C)‡ ‡	ΔTm (°C)
ADF2	F-TTGGAGAACCAGCAGAAAC R-CCATCAAGCTCTCTCTTGAAC	No	No	No	1 (-0.5)	6 (-2.7)	-8.5 → -6.7	-8.1 → -6.7	64.8 65.2	0.4
CYB5	F-GCACGCTACGAAAGAGG R-CAGGATGATCATCTAGAAAAGG	No	No	No	No	1 (1.4)	-9.9 → -7.9	-8.2 → -8.5	64.0 64.0	0.0
iPGAM	F-GGGAGATGTTTCAATTCCAAG R-CCCATTAGAGAAAGATGAGCAG	No	No	No	1 (-0.2)	1 (0.8)	-9.4 → -8.5	-9.6 → -8.2	66.9 66.5	0.4
SCL13	F-AGTCGGTTAGCACGGTTA R-TTCGTGTTCGATTCTTGTT	1 (-1.0)	No	No	No	1 (-0.7)	-8.1 → -7.3	-8.4 → -7.0	62.7 62.7	0.0
TRXL3-3	F-TGGTGTCGTGTTTGTGCAGAG R-GTTGGGTTGTTTCTGGGCATT	No	No	1 (-1.6)	No	No	-8.2 → -6.7	-10.0 → -8.4	71.9 71.6	0.3
VHA-H	F-TGCAAGTGATGATGTTTTGA R-TGCTTGAACAAATGAAGACC	No	No	No	1 (0.7)	1 (0.7)	-8.8 → -7.3	-8.5 → -7.6	64.6 64.8	0.2

‡ ≥2 bp.
‡ ‡ Tm: Melting temperature

Table S2. Log_2FC (fold change) values of six candidate reference genes (RGs) for three different experiments: green vs. red, leaf vs. stem, and drought stress.

Gene	Green vs. red	Leaf vs. stem	Drought stress[a]					
			'Romired'		'Morada de Belchite'		*L. homblei*	
			C vs. DI-1	C vs. DI-2	C vs. DI-1	C vs. DI-2	C vs. DI-1	C vs. DI-2
ADF2	0.298	-0.286	-0.536	-0.677	0.515	-0.234	-0.276	-0.274
CYB5	-0.082	-0.146	-0.175	-0.109	0.231	0.095	0.051	0.068
iPGAM	0.041	-0.241	0.263	0.487	0.226	0.117	0.458	0.451
SCL13	0.173	-0.372	-0.228	-0.018	0.632	0.626	0.591	1.033
TRXL3-3	0.241	-0.459	0.035	0.041	0.090	-0.215	0.239	0.335
VHA-H	-0.208	-0.166	0.152	0.475	0.408	0.210	-0.246	-0.089

[a]C: Control (full irrigation, week 1: 1350 mL, weeks 2-3: 2100 mL/each), DI-1 (Deficit Irrigation 1, week 1: 450 mL, weeks 2-3: 150 mL/each); DI-2 (weeks 1-3: 0 mL).

Table S3. Cq (quantification cycle), efficiency and corrected Cq (CqE) values for six candidate reference genes (RGs) in the three different experiments: green vs. red, leaf vs. stem, and drought stress.

Experiment	Accession	Condition	Biological replicate	Technical replicate	*ADF2*			*CYB5*			*iPGAM*			*SCL13*			*TRXL3-3*			*VHA-H*		
					Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE
Green vs. red	'Begoña'	Green	BR1	TR1	32.270	1.842	28.429	32.370	1.878	29.437	34.763	1.917	32.636	37.223	1.379	17.250	34.519	1.886	31.602	33.500	1.904	31.126
				TR2	33.266	1.877	30.213	32.153	1.925	30.370	36.089	1.859	32.280	37.243	1.383	17.437	34.949	1.795	29.497	32.683	1.891	30.039
			BR2	TR1	32.543	1.857	29.062	31.380	1.925	29.645	37.642	1.701	28.857	36.681	1.446	19.511	35.518	1.834	31.090	31.668	1.944	30.368
				TR2	32.966	1.885	30.147	31.377	1.926	29.662	37.229	1.731	29.463	37.074	1.383	17.340	35.616	1.838	31.266	31.929	1.908	29.755
			BR3	TR1	34.847	1.815	29.969	33.687	1.863	30.231	36.982	1.766	30.346	38.000	1.292	14.065	35.440	1.838	31.111	35.619	1.744	28.584
				TR2	34.236	1.875	31.041	33.554	1.912	31.380	37.961	1.701	29.095	37.658	1.319	15.051	35.446	1.842	31.235	36.024	1.663	26.427
	'Romired'	Red	BR1	TR1	31.431	1.855	28.009	31.302	1.919	29.433	37.311	1.717	29.092	35.869	1.528	21.951	35.119	1.836	30.795	33.386	1.917	31.352
				TR2	31.935	1.873	28.902	31.537	1.904	29.294	37.052	1.740	29.615	34.938	1.607	23.895	35.193	1.842	31.015	33.043	1.852	29.387
			BR2	TR1	33.129	1.853	29.480	32.102	1.905	29.846	36.463	1.822	31.570	36.282	1.494	21.003	36.281	1.744	29.112	31.830	1.928	30.150
				TR2	32.653	1.861	29.258	32.164	1.917	30.188	36.005	1.866	32.395	36.873	1.417	18.528	35.776	1.834	31.314	31.767	1.917	29.817
			BR3	TR1	32.524	1.863	29.188	32.273	1.891	29.657	38.000	1.686	28.637	36.878	1.425	18.832	34.820	1.852	30.961	32.060	1.926	30.322
				TR2	32.609	1.852	29.002	31.925	1.925	30.153	37.499	1.711	29.044	36.781	1.427	18.863	34.710	1.876	31.504	31.570	1.758	25.698
Leaf vs. stem	*L. squarrosa*	Leaf	BR1	TR1	32.692	1.854	29.126	32.737	1.939	31.281	36.590	1.741	29.280	37.643	1.275	13.187	34.192	1.892	31.465	32.251	1.926	30.497
				TR2	32.675	1.875	29.643	32.735	1.927	30.967	36.559	1.745	29.364	37.270	1.314	14.697	34.053	1.889	31.254	32.490	1.893	29.916
			BR2	TR1	32.180	1.840	28.300	32.106	1.937	30.619	36.674	1.742	29.356	35.735	1.493	20.653	34.870	1.848	30.886	33.709	1.938	32.179
				TR2	32.675	1.869	29.479	32.109	1.908	29.921	36.715	1.737	29.263	35.845	1.472	20.009	34.788	1.890	31.939	34.255	1.838	30.078
			BR3	TR1	35.104	1.851	31.182	34.059	1.919	32.026	37.400	1.693	28.396	38.000	1.205	10.207	36.093	1.808	30.850	36.137	1.761	29.497
				TR2	35.675	1.809	30.499	34.067	1.856	30.401	37.040	1.712	28.739	37.347	1.295	13.924	35.891	1.811	30.755	35.808	1.823	31.020
		Stem	BR1	TR1	32.117	1.837	28.172	31.518	1.900	29.197	36.266	1.757	29.491	37.613	1.250	12.130	34.846	1.877	31.642	32.378	1.926	30.606
				TR2	32.559	1.882	29.714	31.783	1.911	29.690	35.513	1.828	30.893	37.629	1.247	11.984	34.663	1.855	30.907	32.031	1.882	29.224
			BR2	TR1	31.135	1.834	27.241	31.983	1.885	29.262	35.856	1.789	30.077	37.740	1.249	12.101	34.059	1.859	30.466	33.123	1.900	30.675
				TR2	31.433	1.841	27.668	31.925	1.875	28.941	35.933	1.824	31.151	37.527	1.280	13.350	34.099	1.871	30.811	32.616	1.868	29.403
			BR3	TR1	31.707	1.848	28.079	30.998	1.898	28.660	37.658	1.663	27.648	34.541	1.586	22.988	35.027	1.842	30.864	35.291	1.845	31.192
				TR2	32.560	1.856	29.041	30.985	1.880	28.211	36.853	1.715	28.683	35.516	1.541	22.169	34.476	1.880	31.395	34.470	1.890	31.654
Drought stress	'Romired'	C	BR1	TR1	35.873	1.680	26.854	34.961	1.843	28.200	NA	NA	NA	36.861	1.414	18.414	36.445	1.703	27.987	36.572	1.663	26.849
				TR2	36.244	1.623	25.336	34.670	1.878	27.626	NA	NA	NA	37.650	1.287	13.687	35.711	1.798	30.226	36.447	1.628	25.624
			BR2	TR1	33.977	1.792	28.605	34.082	1.911	29.887	36.995	1.640	26.401	35.001	1.602	23.787	36.921	1.667	27.211	36.171	1.708	27.945
				TR2	33.790	1.798	28.602	33.615	1.919	29.531	37.029	1.606	25.315	34.437	1.643	24.671	37.074	1.680	27.764	36.116	1.708	27.885
			BR3	TR1	33.104	1.848	29.332	32.691	1.920	29.142	36.498	1.679	27.286	35.284	1.584	23.416	34.093	1.864	30.630	34.974	1.772	28.874
				TR2	33.640	1.809	28.782	33.070	1.902	28.964	36.396	1.674	27.068	33.997	1.639	24.226	34.280	1.824	29.726	35.261	1.790	29.606
		DI-1	BR1	TR1	34.212	1.792	28.800	34.018	1.926	28.230	36.646	1.670	27.125	35.468	1.564	22.886	34.960	1.824	30.303	34.628	1.840	30.468
				TR2	33.548	1.800	28.445	34.050	1.884	28.405	36.681	1.634	26.001	35.623	1.521	21.553	35.225	1.806	30.052	34.724	1.838	30.481
			BR2	TR1	33.801	1.786	28.287	34.228	1.894	28.239	36.258	1.700	27.769	36.118	1.484	20.581	34.577	1.825	30.003	33.771	1.837	29.630
				TR2	33.889	1.795	28.614	33.796	1.896	28.093	36.230	1.686	27.299	35.559	1.526	21.682	34.746	1.815	29.886	33.631	1.853	29.933
			BR3	TR1	32.241	1.820	27.845	33.558	1.922	28.196	35.229	1.753	28.540	34.935	1.595	23.516	33.099	1.880	30.146	33.328	1.912	31.158
				TR2	32.499	1.845	28.728	33.371	1.890	28.117	35.309	1.737	28.125	34.898	1.579	23.000	32.871	1.888	30.148	33.161	1.863	29.760
		DI-2	BR1	TR1	33.976	1.788	28.493	35.002	1.838	28.610	35.237	1.741	28.183	36.858	1.393	17.634	34.116	1.866	30.700	35.282	1.765	28.921
				TR2	34.449	1.773	28.448	35.597	1.816	27.698	35.360	1.747	28.458	36.464	1.455	19.719	34.032	1.822	29.455	35.605	1.740	28.443
			BR2	TR1	37.091	1.550	23.463	38.000	1.583	29.549	36.419	1.674	27.079	38.000	1.260	12.658	37.574	1.637	26.722	37.619	1.554	23.935
				TR2	38.000	1.440	19.978	37.688	1.631	30.452	37.348	1.583	24.744	38.897	1.132	6.956	38.000	1.571	24.766	38.000	1.525	23.131
			BR3	TR1	36.020	1.662	26.388	35.331	1.819	28.305	36.107	1.695	27.474	35.927	1.509	21.334	34.939	1.822	30.237	34.892	1.831	30.458
				TR2	35.972	1.634	25.486	35.512	1.824	28.657	36.086	1.676	26.873	34.966	1.590	23.394	34.960	1.823	30.294	34.937	1.813	29.994

Table S3. Continued.

Experiment	Accession	Condition	Biological replicate	Technical replicate	*ADF2*			*CYB5*			*iPGAM*			*SCL13*			*TRXL3-3*			*VHA-H*		
					Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE	Cq	Efficiency	CqE
Drought stress	'Morada de Belchite'	C	BR1	TR1	33.503	1.865	30.124	34.044	1.915	31.921	34.213	1.849	30.337	36.355	1.458	19.790	34.421	1.831	30.031	34.517	1.845	30.509
				TR2	33.531	1.814	28.801	33.732	1.895	31.108	34.516	1.833	30.162	35.701	1.543	22.355	34.480	1.831	30.098	34.678	1.765	28.428
			BR2	TR1	31.774	1.873	28.773	31.943	1.942	30.585	35.287	1.752	28.538	37.678	1.234	11.411	34.855	1.828	30.327	33.682	1.904	31.286
				TR2	31.949	1.851	28.374	32.096	1.915	30.081	35.749	1.729	28.254	37.519	1.304	14.349	34.592	1.816	29.789	33.499	1.907	31.196
			BR3	TR1	32.685	1.843	28.835	31.757	1.935	30.241	35.078	1.771	28.928	34.752	1.618	24.121	34.693	1.825	30.122	32.952	1.845	29.111
				TR2	33.167	1.797	28.036	31.850	1.911	29.761	35.517	1.732	28.152	34.590	1.620	24.076	34.528	1.850	30.633	32.766	1.890	30.092
		DI-1	BR1	TR1	34.266	1.776	28.389	32.209	1.926	30.454	34.800	1.819	30.024	35.559	1.587	23.688	35.773	1.754	29.000	34.399	1.865	30.926
				TR2	34.236	1.806	29.200	32.680	1.903	30.325	34.775	1.830	30.322	35.482	1.532	21.841	34.715	1.828	30.224	34.484	1.836	30.239
			BR2	TR1	33.858	1.824	29.349	33.905	1.935	32.280	NA	NA	NA	38.359	1.172	8.787	35.449	1.826	30.801	35.380	1.778	29.365
				TR2	34.542	1.781	28.758	33.874	1.909	31.586	NA	NA	NA	37.796	1.284	13.629	35.311	1.820	30.513	35.316	1.754	28.642
			BR3	TR1	33.513	1.820	28.952	34.749	1.879	31.617	35.565	1.763	29.092	36.467	1.433	18.942	35.423	1.813	30.395	34.767	1.853	30.931
				TR2	34.012	1.792	28.618	34.476	1.889	31.627	34.540	1.840	30.389	35.965	1.468	19.912	35.285	1.818	30.437	34.803	1.851	30.928
		DI-2	BR1	TR1	37.362	1.530	22.925	35.449	1.842	31.249	NA	NA	NA	36.825	1.438	19.314	36.969	1.674	27.482	37.542	1.580	24.789
				TR2	37.551	1.500	21.981	35.584	1.822	30.801	NA	NA	NA	36.837	1.378	17.048	35.645	1.788	29.868	36.918	1.638	26.287
			BR2	TR1	35.470	1.655	25.791	33.853	1.885	30.953	35.719	1.692	27.104	36.390	1.422	18.483	34.812	1.829	30.313	34.908	1.827	30.344
				TR2	35.120	1.691	26.605	33.757	1.905	31.390	35.279	1.747	28.388	36.169	1.508	21.438	34.940	1.830	30.463	35.151	1.801	29.847
			BR3	TR1	36.379	1.616	25.177	34.954	1.861	31.326	37.059	1.638	26.369	37.734	1.283	13.559	NA	NA	NA	36.050	1.724	28.326
				TR2	36.511	1.608	25.012	34.516	1.867	31.101	38.000	1.502	22.306	37.819	1.251	12.226	NA	NA	NA	35.908	1.730	28.398
	L. homblei	C	BR1	TR1	32.563	1.839	28.614	31.223	1.870	30.832	35.921	1.732	28.478	36.690	1.429	18.910	35.410	1.811	30.349	33.936	1.877	30.825
				TR2	32.660	1.863	29.307	31.616	1.832	31.531	36.042	1.712	27.943	36.765	1.406	18.070	35.416	1.826	30.768	33.870	1.896	31.254
			BR2	TR1	32.632	1.858	29.157	32.219	1.902	31.838	35.809	1.724	28.145	35.889	1.525	21.835	36.533	1.688	27.604	34.508	1.819	29.789
				TR2	32.818	1.838	28.806	32.299	1.885	31.619	36.123	1.711	27.975	35.755	1.536	22.147	36.562	1.705	28.131	34.161	1.820	29.518
			BR3	TR1	33.075	1.838	29.033	31.440	1.901	30.765	39.088	1.362	17.402	37.840	1.310	14.741	36.577	1.698	27.924	33.436	1.907	31.130
				TR2	32.996	1.857	29.473	31.520	1.891	30.671	39.262	1.354	17.180	37.844	1.284	13.634	36.011	1.733	28.563	33.333	1.821	28.816
		DI-1	BR1	TR1	32.759	1.843	28.901	30.610	1.895	32.156	35.352	1.737	28.158	33.908	1.695	25.822	35.278	1.816	30.374	32.960	1.923	31.102
				TR2	32.583	1.873	29.510	30.585	1.904	31.106	35.400	1.744	28.419	34.495	1.655	25.073	35.153	1.829	30.631	32.658	1.881	29.759
			BR2	TR1	32.014	1.749	25.825	30.909	1.884	31.541	35.030	1.813	30.083	34.189	1.683	25.669	35.364	1.826	30.719	34.299	1.846	30.323
				TR2	31.651	1.862	28.385	30.972	1.875	31.187	34.613	1.826	30.057	34.805	1.601	23.626	35.387	1.815	30.423	33.785	1.898	31.240
			BR3	TR1	32.964	1.830	28.733	30.937	1.881	31.639	33.191	1.881	30.258	35.832	1.529	21.949	35.286	1.818	30.435	35.645	1.724	28.004
				TR2	32.423	1.844	28.627	31.099	1.871	30.639	34.805	1.814	29.917	35.716	1.533	22.016	35.868	1.777	29.761	35.260	1.789	29.588
		DI-2	BR1	TR1	33.761	1.770	27.798	31.543	1.875	30.739	34.203	1.821	29.568	34.128	1.687	25.737	34.250	1.858	30.603	32.374	1.926	30.614
				TR2	33.511	1.784	27.976	31.649	1.834	30.627	34.013	1.839	29.887	33.767	1.703	25.944	34.394	1.828	29.943	32.432	1.878	29.476
			BR2	TR1	34.920	1.704	26.845	32.797	1.867	25.196	NA	NA	NA	33.849	1.728	26.704	35.236	1.844	31.110	33.491	1.892	30.806
				TR2	34.668	1.734	27.522	32.638	1.909	26.612	NA	NA	NA	33.839	1.710	26.193	34.928	1.834	30.561	33.104	1.901	30.673
			BR3	TR1	32.419	1.844	28.616	30.503	1.903	30.496	32.105	1.922	30.257	34.292	1.694	26.091	34.552	1.823	29.932	32.273	1.922	30.429
				TR2	32.296	1.870	29.168	30.375	1.923	30.796	33.586	1.913	31.436	34.447	1.649	24.871	34.699	1.829	30.220	31.813	1.899	29.433

Table S4. Comprehensive stability rankings of six candidate reference genes (RGs) combining geNorm + NormFinder and BestKeeper + ΔCt results for three different experiments: green vs. red, leaf vs. stem, and drought stress.

	Green vs. red		Leaf vs. stem		Drought stress	
Ranking	**geNorm + NormFinder**	**BestKeeper + ΔCt**	**geNorm + NormFinder**	**BestKeeper + ΔCt**	**geNorm + NormFinder**	**BestKeeper + ΔCt**
1	*CYB5 = TRXL3-3*	*CYB5*	*TRXL3-3 = ADF2*	*TRXL3-3*	*TRXL3-3 = ADF2*	*TRXL3-3 = SCL13*
2	*ADF2*	*TRXL3-3 = ADF2*	*CYB5 = VHA-H*	*CYB5 = iPGAM*	*CYB5 = iPGAM*	*iPGAM*
3	*VHA-H*	*SCL13*	*iPGAM*	*ADF2*	*SCL13*	*ADF2*
4	*iPGAM*	*iPGAM*	*SCL13*	*SCL13 = VHA-H*	*VHA-H*	*CYB5*
5	*SCL13*	*VHA-H*	*TRXL3-3 = ADF2*	*TRXL3-3*		*VHA-H*

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 January 2025
DOI 10.3389/fpls.2024.1494339

OPEN ACCESS

EDITED BY
Parimalan Rangan,
Indian Council of Agricultural Research (ICAR), India

REVIEWED BY
Gograj Singh Singh Jat,
Division of Vegetable Science - IARI, India
Akanksha Singh,
Purdue University, United States

*CORRESPONDENCE
Aurora Díaz
adiazb@cita-aragon.es;
adiazb@unizar.es

RECEIVED 10 September 2024
ACCEPTED 23 December 2024
PUBLISHED 15 January 2025

CITATION
Medina-Lozano I, Grimplet J and Díaz A (2025) Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress.
Front. Plant Sci. 15:1494339.
doi: 10.3389/fpls.2024.1494339

Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress

Inés Medina-Lozano[1,2], Jérôme Grimplet[1,2] and Aurora Díaz[1,2]*

[1]Department of Plant Sciences, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain, [2]AgriFood Institute of Aragon – IA2 (CITA-University of Zaragoza), Zaragoza, Spain

Lettuce is a crop particularly vulnerable to drought. A transcriptomic study in the variety 'Romired' and the wild relative *Lactuca homblei* was conducted to understand the increase in anthocyanins (only significant in *L. homblei*) in response to drought previously observed. RNA-seq revealed more differentially expressed genes (DEGs), especially upregulated, in the wild species, in which the most abundant and significant GO terms were involved in regulatory processes (including response to water). Anthocyanin synthesis was triggered in *L. homblei* in response to drought, with 17 genes activated out of the 36 mapped in the phenylpropanoid-flavonoid pathway compared to 7 in 'Romired'. Nineteen candidate DEGs with the strongest change in expression and correlation with both anthocyanin content and drought were selected and validated by qPCR, all being differentially expressed only in the wild species with the two techniques. Their functions were related to anthocyanins and/or stress response and they harboured 404 and 11 polymorphisms in the wild and cultivated species, respectively. Some wild variants had high or moderate predicted impacts on the respective protein function: a transcription factor that responds to abiotic stresses, a heat shock protein involved in stomatal closure, and a phospholipase participating in anthocyanin accumulation under abiotic stress. These genetic variants could explain the differences in the gene expression patterns between the wild (significantly up/downregulated) and the cultivated (no significant changes) species. The diversity of this crop wild relative for anthocyanin-related genes involved in the response to drought could be exploited to improve lettuce resilience against some adverse climate effects.

KEYWORDS

abiotic stress, antioxidants, crop wild relatives, differentially expressed genes, *Lactuca sativa* L., real-time qPCR, RNA-seq, resilience

1 Introduction

Abiotic stress is a major challenge for agriculture, especially in the present scenario of climate change (IPCC, 2021), in which adverse environmental conditions are more and more frequent (FAO, 2021). In particular, drought is one of the most concerning abiotic stresses, affecting both crop yield and quality. Drought stress has important effects on plant growth by affecting diverse physiological and biochemical processes, like cell expansion and photosynthesis due to stomatal closure (Farooq et al., 2009). Apart from biomass production, it also affects nutrient composition and concentration as well as secondary metabolism, depending generally on the stress severity and duration, as well as on plant tolerance (Reddy et al., 2004; Medina-Lozano et al., 2024).

Lettuce (*Lactuca sativa* L.) is one of the most important leafy vegetables worldwide (FAOSTAT, 2021). It provides different health benefits attributed to phenolic compound, vitamin, and fibre contents (Llorach et al., 2008), among others, what contributes to increase its popularity especially with the growing awareness of the impact of diet on health among consumers. Lettuce is mostly composed by water (up to 97%) (Mou, 2005), what makes it highly susceptible to drought (Eriksen et al., 2016). However, controlled deficit irrigation can cause an improvement of its health-promoting properties by increasing the content of some antioxidants (Paim et al., 2020; Medina-Lozano et al., 2024). Among the phenolic compounds present in lettuce, anthocyanins are responsible for red pigmentation of the leaves in semi-red and red varieties. They are known to play crucial roles in human health due to their antioxidant properties (Garcia and Blesso, 2021). It has been described that water stress causes an accumulation of anthocyanins in some fruits, vegetables and oil crops, such as grapes (Ju et al., 2019), strawberries (Rugienius et al., 2021), and purple-stem canola (Chen et al., 2022b). In lettuce, different studies had reported increased levels of either total phenolic compounds under water stress (Zeljković et al., 2023), or anthocyanins in response to other environmental stresses, like UV irradiance (Tsormpatsidis et al., 2008) and low temperatures (Becker et al., 2014). However, anthocyanin response to drought conditions had barely been studied in this crop until recently, when a drought-induced anthocyanin accumulation not only in cultivated lettuce varieties but also in wild relative species, has been discovered (Medina-Lozano et al., 2024).

Lettuce anthocyanin content is very dependent on the genotype. In absence of stress, commercial varieties are the richest, followed by traditional ones and finally by lettuce wild relatives (Medina-Lozano et al., 2021). Interestingly, in all the lettuce-related germplasm studied, the water stress always resulted in an increase of the total anthocyanin content, with the highest accumulation detected in a wild relative species (Medina-Lozano et al., 2024). Crop wild relatives (CWR) are known to be a source of favourable alleles for interesting traits for breeding, like resistance to diseases or tolerance to abiotic stresses (Quezada-Martinez et al., 2021).

Unveiling the molecular mechanisms governing the changes of anthocyanin content in response to water stress could have multiple benefits from a breeding perspective, aiming at enhancing the drought tolerance of the crop and the antioxidant properties of the food product. In lettuce, the great majority of transcriptomic studies related to anthocyanins are focused on the differences between green and red varieties (Moreno-Escamilla et al., 2020; Su et al., 2020). RNA-seq has also been used to study different abiotic stresses in this crop, e.g., high and low temperatures (Park et al., 2020; Chen et al., 2022a), the presence of heavy metals (Xiong et al., 2021), and even drought (Koyama et al., 2021). However, the specific effect of environmental factors on lettuce anthocyanin regulation has been scarcely studied, except in the case of different light conditions (Zhang et al., 2018; Wada et al., 2022).

Nowadays, RNA-seq is the most widely used technology for studying gene expression due to its many advantages. RNA-seq is a precise and sensitive technique that has also a wide range of detection and is highly accurate in terms of quantification (Wang et al., 2009). Despite being a powerful technique, some artefacts may be present in RNA-seq data (Everaert et al., 2017). Therefore, their validation with an independent technique like real-time quantitative PCR (qPCR) is advisable and even necessary when genes are small, have few exons or low levels of expression (Everaert et al., 2017).

Once differentially expressed genes (DEGs) have been identified, the study of polymorphisms in their sequences might provide information about functional and structural effects that could explain the observed variation for the trait of interest. However, the elucidation of these effects through experimental approaches is usually time and labour consuming and, in many cases, leads to dead ends. That is why the development and use of computational prediction tools as a first approach have experienced a boom in the last few years as they are able to provide increasingly more accurate information to assess phenotypic effects (Yazar and Özbek, 2021).

Metabolite-mediated drought adaptation is an emerging subject that has revealed the importance of some primary metabolites, such as sugars, small peptides, and amino acids, among others, in plant response, either acting as signal factors or as protectors (Zhang et al., 2024). Less is known about the participation of secondary metabolites (e.g., anthocyanins) in plant response to drought, beyond their antioxidant activity like scavengers of reactive oxygen species (ROS) (Naing and Kim, 2021). In this work, we have carried out transcriptomic analyses via RNA-seq and real-time qPCR in a red lettuce variety and a wild relative species that

Abbreviations: ABA, abscisic acid; C, control; cDNA, complementary deoxyribonucleic acid; CHS, chalcone synthase; CITA, Agrifood Research and Technology Centre of Aragón; CPM, counts per million; CWR, crop wild relatives; DEG, differentially expressed gene; DI, deficit irrigation; EBG, early biosynthesis gene; FC, fold change; FDR, false discovery rate; GO, gene ontology; HSP, heat shock protein; Indel, insertion-deletion; LBG, late biosynthesis gene; MNP, multiple nucleotide polymorphisms; MYB, myeloblastosis; NAC, NAM (no apical meristem), ATAF (*Arabidopsis thaliana* activating factor), and CUC (cup-shaped cotyledon); PLIP, phospholipid-inositol phosphatase; PTFE, polytetrafluoroethylene; qPCR, quantitative polymerase chain reaction; RNA-seq, ribonucleic acid sequencing; ROS, reactive oxygen species; RTL3, ribonuclease III-like protein 3; siRNA, small interfering RNA; SNP, single nucleotide polymorphism; spp., species; TF, transcription factor; Vs., versus; WGCNA, weighted gene co-expression network analysis.

experienced a raise in anthocyanin content as a response to drought stress (Medina-Lozano et al., 2024). In addition, *in silico* predictions of the effects of polymorphisms in DEGs could potentially explain the observed differences between the two species in anthocyanin content in plants subject to water stress. The genetic knowledge of this response is key to obtaining new lettuce varieties with both enhanced drought tolerance and health-promoting properties, at the same time that water resources destined to irrigation could be cut down.

2 Materials and methods

2.1 Plant material

Two different accessions of the genus *Lactuca* were included in this study: a commercial variety, the red-leaf lettuce 'Romired', and a wild relative species, *Lactuca homblei* De Wild. They were selected from a previous drought stress experiment in which two irrigation regimes, control (C, week 1: 1350 mL, weeks 2-3: 2100 mL/each) and deficit irrigation (DI, weeks 1-3: 0 mL), were tested in two consecutive years (Medina-Lozano et al., 2024). Three biological replicates for the two accessions in each of the two conditions (C and DI) from the experiment carried out in winter 2020/2021 were used to proceed with the transcriptomic studies.

2.2 RNA extraction and sequencing, data processing and DEG identification

Total RNA extraction from lyophilized samples coming from 12 samples (2 accessions x 2 irrigation regimes x 3 biological replicates) was performed using the NZY Total RNA Isolation kit (NZYtech Lda.-Genes and Enzymes, Lisbon, Portugal) as described before (Medina-Lozano et al., 2023). RNA was treated with DNase using the TURBO DNA-free™ kit (Invitrogen, Waltham, MA, USA), following the manufacturers' instructions. RNA quantity and purity were assessed with the Eukaryotic Total RNA Nanobioanalyzer Assay in a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

The obtained RNA samples from 'Romired' and *L. homblei* were used to perform the RNA-seq. They were processed to build a total of 12 strand-specific cDNA libraries. Sequencing of the libraries was performed in both directions with a NovaSeq 6000 S1 instrument (Illumina, San Diego, CA, USA) using the TruSeq Stranded mRNA protocol (Illumina) to obtain between 36 and 111 strand-specific pair-end reads of 100 base pair (bp) lengths per sample. Sequencing was carried out at the National Centre for Genomic Regulation (CNAG-CRG, Barcelona, Spain).

Sequences were analysed using the Galaxy tool (The Galaxy Community, 2022). Adapter sequences were removed by processing the reads from the 12 individual datasets using Trimmomatic (Galaxy version 0.38.1) (Bolger et al., 2014). RNA-seq data alignment to the lettuce reference genome *Lactuca sativa* 'Salinas' v8 (Reyes-Chin-Wo et al., 2017) was performed using HISAT2 (Galaxy version 2.2.1+galaxy0) (Kim et al., 2015), with a maximum intron length set at 20,000 bp. The Picard tools (http://broadinstitute.github.io/picard) MarkDuplicates (Galaxy version 2.18.2.2) and FixMateInformation (Galaxy version 2.18.2.1) were used to filter out the optical duplicates and to mate-pairs, respectively. featureCounts (Galaxy version 2.0.1+galaxy2) (Liao et al., 2013) was used to generate read counts using the gene annotation available in the literature (Reyes-Chin-Wo et al., 2017).

Analysis of differential gene expression between treatments (C and DI) within each of the two accessions was conducted using edgeR (Galaxy version 3.36.0+galaxy0) (Robinson et al., 2009). Genes were considered to be differentially expressed when values of $|\log_2(\text{FC, fold change})|>1$ and FDR (False Discovery Rate)<0.05 (adjusted *p*-value via the Benjamini-Hochberg method).

2.3 Structural and functional analysis of the DEGs

Venn diagrams were performed with DEG datasets using the R stats package VennDiagram (https://CRAN.R-project.org/package=VennDiagram). GO (Gene Ontology) enrichment analyses were carried out using the tool GOEnrichment from Galaxy platform (Galaxy version 2.0.1) (Faria, 2017), with *p*-value cut-off < 0.05 and using Benjamini-Hochberg multiple test correction. Enriched GO terms involving three categories, biological processes, cellular components, and molecular functions, were evaluated. GO terms of the DEGs were obtained from predicted data using information available in the literature (Reyes-Chin-Wo et al., 2017). Heatmaps were constructed using gplots (https://CRAN.R-project.org/package=gplots) and ggplot2 (Wickham, 2009) R stats packages.

2.4 Selection of DEGs

The selection of genes for expression data validation was based on different criteria. First, genes were filtered out for values of $|\log_2(\text{FC})|>4$ and FDR<0.05 (substantial increase or decrease in expression levels). Among them, those exhibiting high and significant correlation (both positive and negative) with, first, anthocyanin content and, second, drought stress treatment, were selected. Finally, gene functions related to both anthocyanin content and/or response to different stresses were also taken into account for the selection of a total of 19 DEGs.

Correlations between gene expression and both anthocyanin content and treatments were established through weighted gene co-expression network analysis (WGCNA), which was conducted using the R stats package WGCNA (Langfelder and Horvath, 2008). Normalised RNA-seq data of all genes were used for the WGCNA, except for those with a very low expression among the no DEGs (i.e., less than 5 reads per sample in the three biological samples of each group (C and DI)). Data from both species, *L. sativa* (cultivated lettuce 'Romired') and *L. homblei*, were analysed separately.

2.5 DEG validation using real-time qPCR

Total RNA was extracted from each of the 12 samples described above. Subsequently, mRNA was purified using the Dynabeads mRNA DIRECT™ kit (Invitrogen) and cDNA was synthesized using the NZY M-MuLV First-Strand cDNA Synthesis, separate oligos kit (NZYTech) as described before (Medina-Lozano et al., 2023).

Specific pairs of primers for each of the 19 selected DEGs (Supplementary Table S1) were designed using OLIGO software version 6.45 (Cascade, CO, USA) from a consensus sequence of the two species under study, *L. sativa* and *L. homblei*, excluding any ambiguity in the sequences. Real-time qPCR reactions were performed on a StepOnePlus™ System (Applied Biosystems, Waltham, MA, USA) with two technical replicates per each of the three biological replicates. Each reaction contained 1 µL of 1:5 diluted cDNA, 0.40 µM of forward and reverse primers (Integrated DNA Technologies, IDT, Coralville, Iowa, USA), and 1x NZYSupreme qPCR Green Master Mix, ROX plus (NZYTech) in a final volume of 11 µL. The amplification conditions were: 2 min at 95°C, 40 cycles of 5 s at 95°C, 15 s at 52-66°C (Supplementary Table S1) and 30 s at 72°C, followed by the melting curve analysis that ranged from 72°C to 95°C with 0.3°C increment per cycle to verify that a single product was amplified. Non-template controls were included to ensure that contamination with genomic DNA had not occurred.

TRXL3-3 was used as reference gene to normalise qPCR data (Medina-Lozano et al., 2023). Relative expression levels were obtained using the Pfaffl method (Pfaffl, 2001) with some modifications: arithmetic instead of geometric mean was calculated due to the presence of zero values in the raw data (either genes completely shut down as a consequence of the DI or the other way round, unexpressed genes in C conditions that were activated with the DI). This explains values different from 1 in C samples and why they have been represented separately from the DI data in qPCR results.

Student *t*-test was used to assess whether the differences between the means from the qPCR expression data of samples under C and DI conditions were statistically significant. Data transformations ($1/(1+x)^2$ or $1/\sqrt{(x+1)}$) were applied when needed to achieve a normal distribution. Alternatively, Wilcoxon test was used with non-normally distributed data. Statistical analyses were conducted using the software JMP v5.1.2 for Windows (SAS Institute Inc. Cary, NC).

2.6 Polymorphism search, annotation, and effect prediction in the DEGs

Detection of polymorphisms was carried out using the sequences of the 12 samples aligned to the lettuce reference genome (Reyes-Chin-Wo et al., 2017) and processed as explained in subsection 2.2. Firstly, variant calling was performed using FreeBayes package (Galaxy Version 1.3.6+galaxy0) (Tange, 2011; Garrison and Marth, 2012) from Galaxy platform. Then, VCFfilter (Galaxy Version 1.0.0_rc3+galaxy3) (Garrison, 2015) was used to remove polymorphisms with a total read depth at the locus < 10, QUAL < 20, and number of alternative alleles in called genotypes > 0. In addition, those polymorphic sites exhibiting different genotypes among the total number of samples within accessions and/or more than two different genotypes in comparison with the reference genome in more than 70% of the cases, were filtered out with Excel. Any possible ambiguous polymorphism was also eliminated.

The effect of each polymorphism was annotated and predicted using the SnpEff eff tool (Galaxy Version 4.3+T.galaxy1) and a snpEff database created using the SnpEff build tool (Galaxy Version 4.3 +T.galaxy4) (Cingolani et al., 2012) from the annotation dataset and the FASTA file of *L. sativa* 'Salinas' v8 (Reyes-Chin-Wo et al., 2017).

3 Results

3.1 Transcriptome analysis

To investigate the involvement of anthocyanins at molecular level in the response mechanism to drought stress of *Lactuca* spp., a transcriptomic analysis via RNA-seq was performed using plants belonging to the CWR *L. homblei* and to the red commercial lettuce variety 'Romired' coming from a previous experiment carried out in winter 2020/2021 (Medina-Lozano et al., 2024). Samples of both accessions showed an accumulation of anthocyanins under DI in comparison to C conditions, though the differences only resulted statistically significant in the case of the wild species *L. homblei* (Medina-Lozano et al., 2024). In particular, three different anthocyanins were identified: cyanidin 3-*O*-(6'-*O*-malonylglucoside) was the predominant one and was detected in both accessions and treatments; peonidin 3-*O*-glucoside appeared under both treatments in the commercial variety, but only under DI in the CWR; and cyanidin 3-(6''-acetylglucoside), exclusively identified under DI conditions in the commercial variety.

After processing the data from *L. homblei*, the clean reads ranged from 40.76 to 51.75 Gb and the percentage of uniquely mapped sequences to the reference genome ranged from 32.39% to 37.40%. In the case of 'Romired', the clean reads ranged from 35.43 to 110.62 Gb and the uniquely mapped sequences from 81.04% to 84.81% (Table 1). RNA-seq data from both accessions were aligned to the lettuce reference genome *L. sativa* 'Salinas' v8 (Reyes-Chin-Wo et al., 2017). However, *L. homblei* belongs to the tertiary lettuce gene pool (PGR (Plant Genetic Resources) Lettuce, https://www.pgrportal.nl/en/lettuce-geneticresources-portal.htm), so it is quite distant from *L. sativa*, what might explain its lower values in terms of uniquely mapped sequences.

3.2 Identification and analysis of DEGs under drought stress conditions

A total of 6,179 DEGs were identified when *L. homblei* plants under C and DI treatments were compared (3,113 upregulated and 3,066 downregulated genes), whereas a total of 5,329 DEGs were obtained in 'Romired' plants for the same treatments (1,747

10.3389/fpls.2024.1494339

TABLE 1 Statistical summary of RNA-sequencing data.

Group[a]	Sample	Raw reads	Clean reads	Mapped reads	Mapping rate (%)	GC (%)
L. homblei C	*L. homblei* 1	51,853,084	51,749,303	19,128,708	36.96	44.5
	L. homblei 2	46,166,391	46,074,390	17,230,542	37.40	45.5
	L. homblei 3	42,697,894	42,615,564	15,099,405	35.43	43.5
L. homblei DI	*L. homblei* 4	43,141,038	43,063,549	15,173,600	35.24	45.0
	L. homblei 5	40,836,180	40,756,877	13,200,520	32.39	45.0
	L. homblei 6	46,624,421	46,532,339	16,185,998	34.78	44.0
'Romired' C	'Romired' 1	35,523,451	35,428,117	29,244,585	82.55	41.5
	'Romired' 2	101,124,691	100,936,321	85,092,426	84.30	44.5
	'Romired' 3	110,827,310	110,624,887	93,466,553	84.49	45.0
'Romired' DI	'Romired' 4	94,626,862	94,435,273	80,092,100	84.81	45.0
	'Romired' 5	52,365,850	52,260,702	42,353,540	81.04	45.0
	'Romired' 6	41,007,770	40,932,006	34,429,220	84.11	45.0

[a]C, control; DI, deficit irrigation.

upregulated and 3,582 downregulated genes) (Figure 1A). A total of 2,272 DEGs were common to both accessions: 847 genes were upregulated, 1,347 downregulated, and 78 exhibited an opposite behaviour in the two accessions (Figure 1B). Attending to the differences, the CWR showed a total number of DEGs higher than the cultivated species (42.3% of exclusive DEGs in *L. homblei* vs. 33.1% in 'Romired'). The same happened in the case of the upregulated genes, where the disparity was the largest, a 56.5% of the DEGs was exclusively upregulated in *L. homblei*, which was more than twice the upregulated DEGs only in 'Romired' (22.4%). In the case of the downregulated genes, we observed the opposite, the number was higher in the cultivated species than in the wild relative (42.2% vs. 32.4%, respectively) (Figure 1B).

To deeply explore the DEG functions in the drought response of *Lactuca* spp., analyses of GO enrichment were conducted using the GO annotations found in Reyes-Chin-Wo et al. (2017). The three main GO categories, biological processes, cellular components and molecular functions, were studied within the upregulated and downregulated genes (Figure 2). Within the upregulated genes, the number of enriched GO terms in biological processes was higher in *L. homblei* than in 'Romired' (Figure 2A). In particular, an important number of *L. homblei* DEGs belonged to significantly

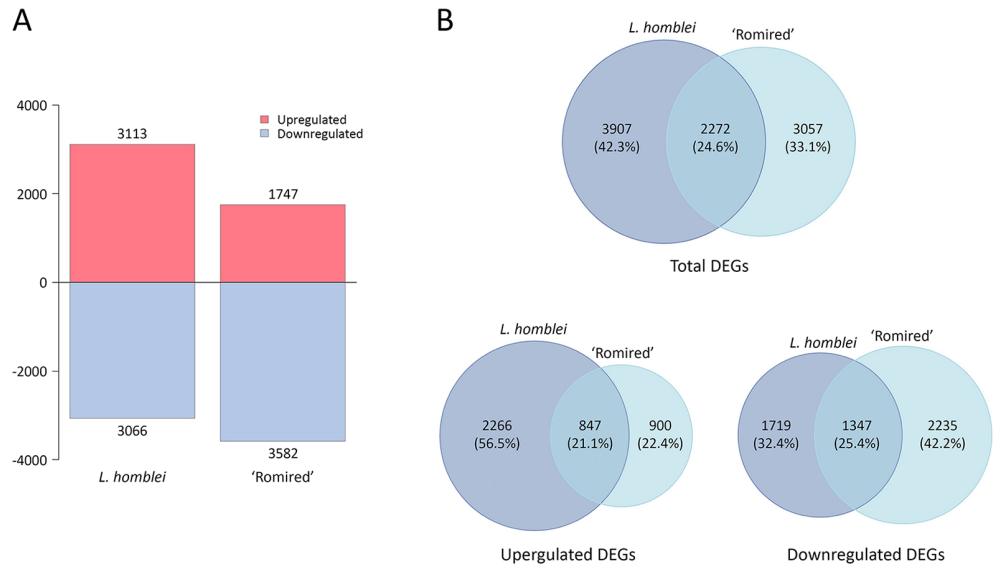


FIGURE 1
Differentially expressed genes (DEGs) in *Lactuca* spp. in response to drought stress. **(A)** Number of upregulated and downregulated DEGs in *L. homblei* and 'Romired'. **(B)** Venn diagrams showing the common and exclusive number of genes within the total, upregulated, and downregulated DEGs.

10.3389/fpls.2024.1494339

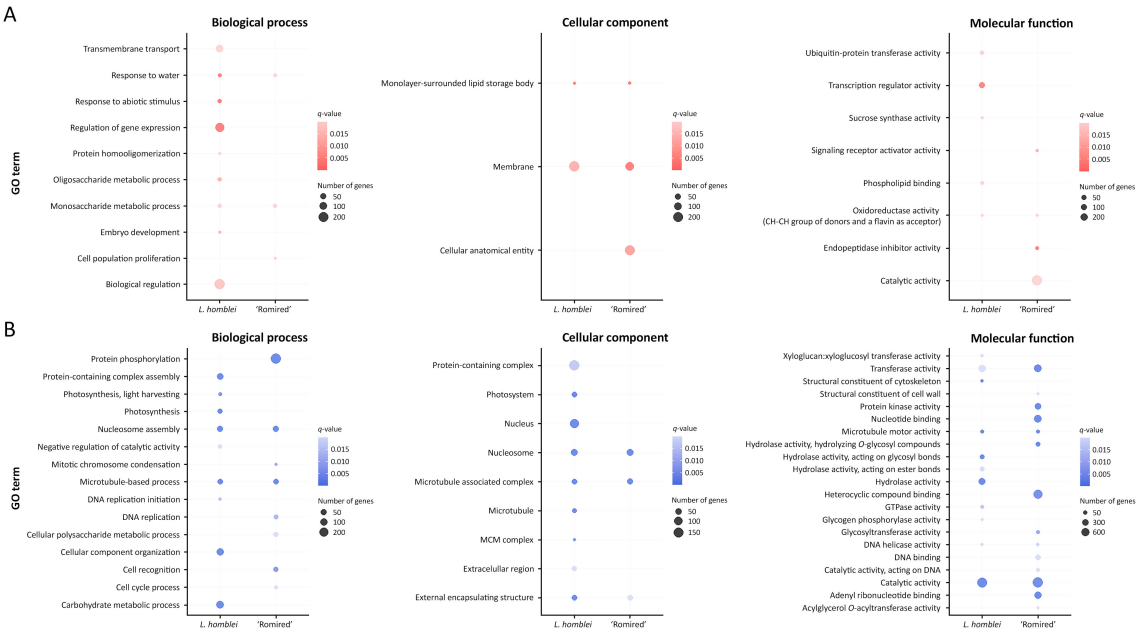


FIGURE 2
Enriched GO terms of *L. homblei* and 'Romired' within the **(A)** upregulated and **(B)** downregulated differentially expressed genes (DEGs) identified in a drought stress experiment for the three main GO categories: biological process, cellular component, and molecular function.

enriched GOs that were involved in transmembrane transport and different metabolic processes, though the most significantly upregulated DEGs were those in enriched GOs related to gene expression regulation and response to abiotic stimulus, water included. The response to water resulted to be also among the enriched GO terms in 'Romired', but the number of genes, and especially the significance level, were lower than in the CWR *L. homblei*. In the case of cellular components (Figure 2A), a similar number of genes were part of the GO terms membranes and lipid storage bodies in both species, with a higher significance in 'Romired'. In addition, the GO term cellular anatomical entity was enriched exclusively in the cultivated accession. Finally, for the molecular function category (Figure 2A), the most significantly enriched term was the transcription regulator activity in the CWR *L. homblei*, and the endopeptidase inhibitor activity in the commercial lettuce 'Romired'. Interestingly, anthocyanins (among other flavins) could be involved in the oxidoreductase activity in which a flavin group acts as acceptor, being actually the only enriched GO common to both species.

Within the downregulated DEGs of biological processes (Figure 2B), genes involved in the carbohydrate metabolism were the most represented in *L. homblei*, while in 'Romired' were those implied in protein phosphorylation, that in fact, appeared only in this accession. However, several enriched processes, as well as their significance levels, were common or similar in both species, such as those related to cellular division and multiplication (DNA replication, nucleosome assembly, and microtubule-based processes). In the cellular component category (Figure 2B), the enriched GO terms found in 'Romired' appeared also enriched in *L. homblei*: nucleosome, microtubule associated complex, and external encapsulating structure, with similar significance and number of genes, except for the external encapsulating structure that resulted more significant in *L. homblei*. In fact, many more GO terms were enriched in *L. homblei*, with the nucleus and the protein-containing complex being the most represented ones. On the contrary, we found many more downregulated DEGs with enriched GO, as well as more terms and with a higher significance, in the cultivated ('Romired') than in the wild species (*L. homblei*) in the molecular function category (Figure 2B).

Some DEGs were assigned to more than one GO term, either because a gene can participate in different biological processes and molecular functions and be part of different cellular components or because GO is loosely hierarchical, with genes belonging to both 'parent' and 'child' terms. Thus, counting genes only once within each category, we obtained that the number of DEGs with enriched GO terms was very similar between *L. homblei* and 'Romired' within the upregulated genes (746 vs. 794, respectively), while it was lower in the CWR *L. homblei* than in the commercial variety 'Romired' within the downregulated ones (1,591 vs. 1978, respectively). Even so, the percentages of common DEGs in the two species was considerably lower in the case of upregulated genes than in downregulated: 3.23%, 18.25%, and 1.14% vs. 12.81%, 22.98%, and 23.00% in biological processes, cellular components and molecular functions, respectively.

Two heatmaps constructed using normalised expression data of DEGs in *L. homblei* and 'Romired' confirmed the effect of the drought stress treatment in *Lactuca* spp. plants in terms of gene regulation (Figure 3). A hierarchical clustering conducted with all the DEGs allowed us to identify two separate groups in the two accessions, as expected, the upregulated and the downregulated ones. This clustering also divided clearly the two conditions (C and DI) in both species, what was even more evident for *L. homblei*

(Figure 3A). In addition, both heatmaps showed again that the number of upregulated genes under water deficit was clearly higher in the CWR *L. homblei* (Figure 3A), whereas those downregulated were more numerous in the commercial variety 'Romired' (Figure 3B). These results show that the wild species was activating more mechanisms in response to drought stress. Figure 3 also shows two heatmaps constructed using the data of the anthocyanin content variation as a consequence of the drought stress for *L. homblei* and 'Romired' (Medina-Lozano et al., 2024), and the data of the two treatments themselves (C and DI). Content of all detected anthocyanins was higher under DI treatment than in C conditions in both *Lactuca* spp. Similar to what happened with the upregulated genes, the accumulation of anthocyanins in response to water stress was higher (and only significant) in *L. homblei* (Figure 3A). Actually, in *L. homblei* all the DI replicates showed a higher content than the C replicates. This was especially remarkable in the case of peonidin 3-*O*-glucoside, which was present under DI conditions and in the 3 biological replicates, but not under C conditions (Figure 3A). However, differences were not so clear (and not significant) between DI and C replicates in 'Romired', despite mean anthocyanin content being higher in DI than in C conditions, as commented above (Figure 3B). Even though, one of the anthocyanins was also identified exclusively under DI conditions in 'Romired', as observed in *L. homblei*, but in this case, it was cyanidin 3-(6''-acetylglucoside) (Figure 3B).

Nevertheless, the molecular mechanisms underlying the anthocyanin accumulation as a consequence of water deficiency have been barely studied in lettuce, unlike in other crops like grapevine (Castellarin et al., 2007) or canola (Chen et al., 2022b). To gain a more comprehensive understanding of the process in *Lactuca* spp., we mapped the expression profiles of the DEGs identified in the drought experiment which participate in the biosynthesis pathway of the detected anthocyanins (the general phenylpropanoid pathway and the flavonoid pathway, this last one leading specifically to the anthocyanin biosynthesis) (Figure 4). We found a total of 36 DEGs involved in the pathway in either *L. homblei*, 'Romired' or both. Different expression profiles were observed between both species. Our results confirmed that, in these routes, more DEGs were activated in the CWR *L. homblei* than in the commercial variety 'Romired', 17 vs. 7 upregulated genes, respectively, what was concordant with the higher accumulation of anthocyanins in the wild relative (Medina-Lozano et al., 2024). Not all the isoforms of the genes coding for the enzymes catalysing each step were upregulated under DI. The activation happened mainly in the first steps of the pathway, that is, in the early biosynthesis genes (EBGs), especially at the beginning of anthocyanin-specific route (flavonoid pathway). This becomes glaringly obvious in the first step which is catalysed by the chalcone synthase (CHS), whose gene isoforms are all strongly and significantly upregulated only in *L. homblei* (triggering of the anthocyanin synthesis in the wild species). This pattern is not so obvious in the preceding genes from the general phenylpropanoid pathway as they participate in the biosynthesis of many other compounds apart from anthocyanins. Furthermore, the late biosynthesis genes (LBGs) were mostly upregulated in the CWR *L. homblei* but not in all the isoforms as observed in *CHS* (EBG), except in the last step which leads to the synthesis of the specific major anthocyanin (cyanidin 3-*O*-(6'-*O*-malonylglucoside)) where most of the genes coding for the isoforms were significantly activated in *L. homblei*. The final steps to produce the two minor anthocyanins, cyanidin 3-(6''-acetylglucoside) and peonidin 3-*O*-glucoside, are not clearly described in the literature. They might be catalysed by some acetyltransferases and *O*-methyltransferases, respectively, as suggested by Ino and Yamaguchi (1993) and Hugueney et al. (2009), respectively. It is possible that the genes coding for these enzymes were activated under drought as those anthocyanins were detected in 'Romired' and *L. homblei*, respectively, only under stress conditions though they have not been characterised in *L. sativa* yet.

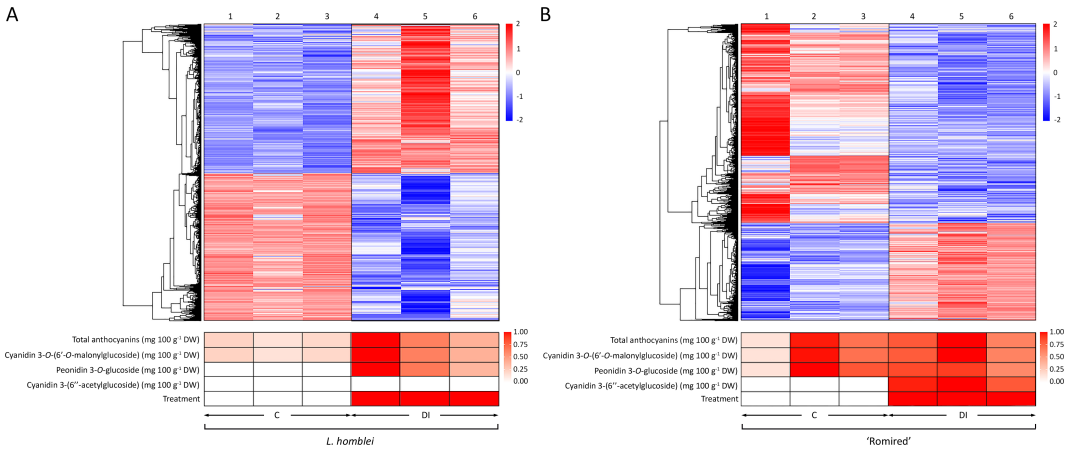


FIGURE 3
Heatmap representation of hierarchical analysis of the expression data from the differentially expressed genes (DEGs), as well as of total and individual anthocyanin content and treatment in **(A)** *L. homblei* and **(B)** 'Romired' under control (C) and deficit irrigation (DI) conditions. Numbers 1-3 and 4-6 show the biological replicates under C and DI, respectively. Phenotypic heatmaps represent scaled data from 0.2 to 1 within each compound, except for minor anthocyanins (peonidin 3-*O*-glucoside in *L. homblei* and cyanidin 3-(6''-acetylglucoside) in 'Romired') where data, as well as treatments, were scaled from 0 to 1.

10.3389/fpls.2024.1494339

3.3 Selection of candidate genes among the DEGs

Selection of anthocyanin-related genes potentially involved in the response to drought stress was based on different criteria. First, we searched for important changes in the expression levels. Second, we selected DEGs with high (positive and negative) and significant values of correlation with both anthocyanin content and drought stress treatment, obtained through a WGCNA. WGCNA allows to identify genes correlated with certain traits (anthocyanins and irrigation treatment in our case) to reveal putative genes with particular interest (Horvath and Dong, 2008). Lastly, we paid attention to gene function, so that DEGs were related to stress and/or anthocyanin content. Finally, 19 genes were selected for validation through real-time qPCR. Remarkably, genes meeting all these criteria resulted to be differentially expressed exclusively in *L. homblei*.

L. homblei $|\log_2(FC)|$ values between C and DI ranged from 4.03 to 6.23 (Table 2). Both up- and downregulated genes were included in the selection. The higher accumulation of anthocyanins and/or the activation of stress response may result either from the upregulation of activators or from the downregulation of repressors. This was also observed in a previous study that characterised four genes related to anthocyanin content in lettuce (Su et al., 2020). From the WGCNA results, we obtained absolute correlation values with anthocyanins ranging from 0.81 to 0.93, and with treatment, from 0.86 to 0.99 (Table 2). Both positive and negative correlations were also considered here. By contrast, in the case of 'Romired', we found 5 out of the 19 genes showing a significant correlation with the anthocyanin content, but none of them exhibited a significant change of expression level nor a significant correlation with the stress treatment (data not shown).

3.4 Validations of candidate genes by qPCR

The expression data of the 19 selected genes obtained from the RNA-seq analysis were validated by real-time qPCR. In *L. homblei*, the 13 downregulated and the 6 upregulated genes according to the RNA-seq analysis showed concordant expression profiles with the qPCR results (Figure 5). Significant, and very significant differences were observed between C and DI treatments for seven and two genes, respectively, according to qPCR data (Figure 5). In the case of 'Romired', the selected genes did not show any differential expression in the RNA-seq analysis, as mentioned above, nor by qPCR. Even so, the expression of 15 out of the 19 genes followed the same profile using the two different techniques (Supplementary Figure S1). Therefore, we were able to confirm the reliability of the results from the RNA-seq analysis.

The expression profiles of the selected DEGs obtained with both techniques (RNA-seq and qPCR) were also analysed by hierarchical clustering both using the mean values (Figure 6) and all data (Supplementary Figure S2). According to the RNA-seq data, two clearly differentiated expression patterns could be observed for *L. homblei*: the expression levels were noticeably lower in C than in DI conditions in the case of upregulated genes, and vice versa in the case of downregulated genes (Figure 6A; Supplementary Figure S2A). In contrast, for 'Romired' the differences in expression were not clear in most of the selected genes (Figure 6A; Supplementary Figure S2A), as was expected since they did not result to be differentially expressed in the RNA-seq analysis. Comparable patterns, especially for *L. homblei* samples, were observed when qPCR data were represented (Figure 6B; Supplementary Figure S2B).

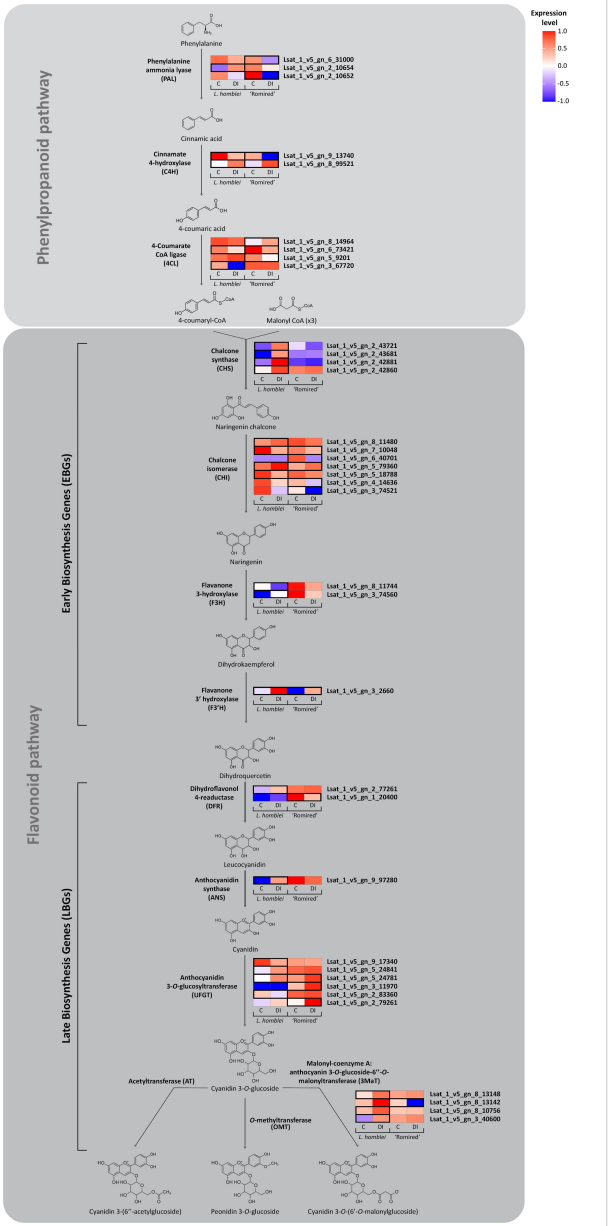


FIGURE 4
Simplified pathway for the anthocyanin biosynthesis of differentially expressed genes (DEGs) detected in *L. homblei* and/or 'Romired' under control (C) and deficit irrigation (DI) conditions. Heatmaps represent the expression data scaled from -1 to 1 for each isoform of the DEGs (same names as the enzymes that catalyse each step) identified in the RNA-seq analysis. Black boxes indicate the accession in which the genes were differentially expressed ($|\log_2(FC)|>1.06$).

TABLE 2 Gene product, regulation and correlation with total anthocyanin content and treatment of the 19 differentially expressed genes (DEGs) selected.

					Total anthocyanins		Irrigation treatment	
Gene ID	Gene product	Regulation	$Log_2(FC)$[a]	FDR[b]	Correlation coefficient	*p* value	Correlation coefficient	*p* value
Lsat_1_v5_gn_1_21441	Subtilisin-like protease SBT3	Downregulated	-4.03	3.33E-04	-0.81	0.05	-0.96	2.85E-03
Lsat_1_v5_gn_1_50480	Haloacid dehalogenase (HAD)-like hydrolase superfamily protein	Downregulated	-4.28	3.33E-04	-0.93	0.01	-0.86	0.03
Lsat_1_v5_gn_1_109200	DNA damage-repair/toleration protein DRT100	Downregulated	-4.44	2.27E-03	-0.82	0.05	-0.91	0.01
Lsat_1_v5_gn_1_127541	14 kDa proline-rich protein DC2.15	Downregulated	-5.14	2.53E-03	-0.93	0.01	-0.91	0.01
Lsat_1_v5_gn_2_15680	GDSL esterase/lipase	Downregulated	-4.81	0.04	-0.93	0.01	-0.89	0.02
Lsat_1_v5_gn_2_43400	Probable pectate lyase 8	Downregulated	-5.52	2.25E-03	-0.87	0.02	-0.99	1.75E-04
Lsat_1_v5_gn_2_47181	Protein ECERIFERUM 26	Downregulated	-5.16	0.03	-0.92	0.01	-0.91	0.01
Lsat_1_v5_gn_2_90361	Gibberellin-regulated protein 6	Downregulated	-4.29	0.03	-0.93	0.01	-0.95	3.94E-03
Lsat_1_v5_gn_2_116640	Heat shock cognate 70 kDa protein 2	Upregulated	5.53	6.34E-06	0.84	0.03	0.97	1.06E-03
Lsat_1_v5_gn_3_1101	Zinc finger protein ZAT1	Downregulated	-4.75	1.31E-03	-0.90	0.01	-0.98	5.92E-04
Lsat_1_v5_gn_3_20640	PRA1 family protein E	Upregulated	4.18	1.18E-03	0.82	0.04	0.98	8.11E-04
Lsat_1_v5_gn_5_7401	NAC transcription factor 56	Upregulated	4.22	7.39E-04	0.82	0.05	0.94	0.01
Lsat_1_v5_gn_5_10141	Protein MHF1 homolog	Downregulated	-4.66	0.04	-0.91	0.01	-0.86	0.03
Lsat_1_v5_gn_5_26000	B-box zinc finger protein 21-like	Upregulated	4.34	2.38E-05	0.82	0.05	0.93	0.01
Lsat_1_v5_gn_6_67540	Type I inositol polyphosphate 5-phosphatase 2	Downregulated	-4.32	7.53E-07	-0.85	0.03	-0.93	0.01
Lsat_1_v5_gn_7_92980	Ribonuclease III-like protein RTL3	Downregulated	-6.23	3.97E-03	-0.92	0.01	-0.93	0.01
Lsat_1_v5_gn_8_157561	Transcription factor MYC/MYB N-terminal domain-containing protein	Downregulated	-4.46	4.60E-04	-0.84	0.04	-0.94	4.91E-03
Lsat_1_v5_gn_8_165301	Phospholipase A1 phospholipid-inositol phosphatase PLIP2	Upregulated	4.28	4.73E-05	0.84	0.04	0.96	1.94E-03
Lsat_1_v5_gn_9_80621	Amino acid permease 6	Upregulated	4.17	2.91E-03	0.82	0.04	0.96	2.81E-03

[a]FC, fold change.
[b]FDR, False Discovery Rate.

10.3389/fpls.2024.1494339

10.3389/fpls.2024.1494339

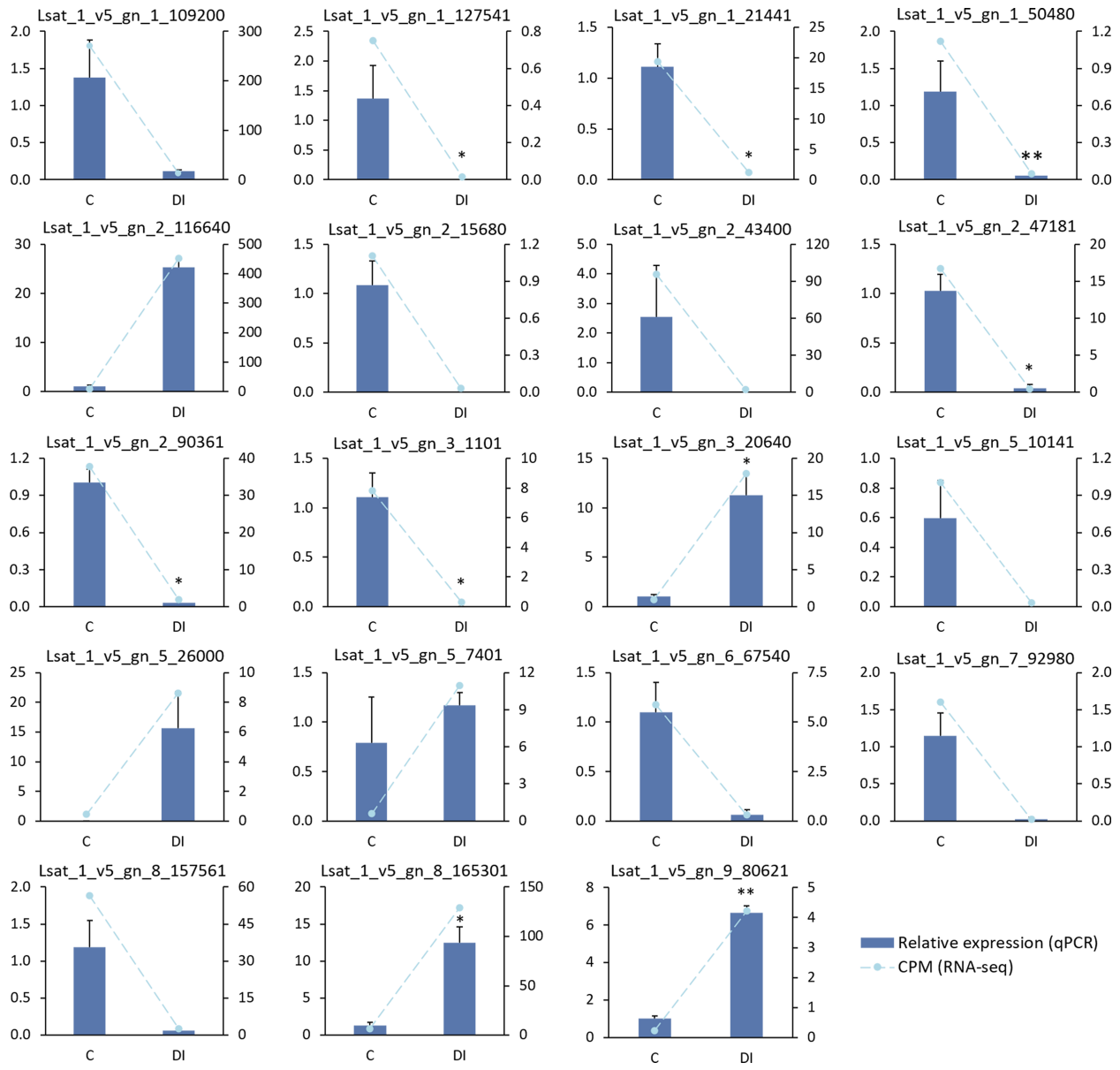


FIGURE 5
Expression data obtained by qPCR (relative expression) and by RNA-seq (CPM, counts per million) of 19 selected genes in the wild species *L. homblei* under control (C) and deficit irrigation (DI) conditions. Bars represent standard error of the total (n=3). *p<0.05, **p<0.01. Transformations were applied to achieve normal distribution in qPCR data in the following cases: $1/(1+x)^2$ to Lsat_1_v5_gn_1_109200, Lsat_1_v5_gn_1_127541, Lsat_1_v5_gn_1_50480, Lsat_1_v5_gn_2_116640, Lsat_1_v5_gn_2_43400, and Lsat_1_v5_gn_3_20640; and $1/\sqrt{(x+1)}$ to Lsat_1_v5_gn_1_21441. Wilcoxon test was used with non-normally distributed qPCR data of Lsat_1_v5_gn_5_10141 and Lsat_1_v5_gn_5_26000.

3.5 Putative function of validated candidate genes

Turning the attention to gene function, it could be confirmed that those 19 DEGs with large changes in expression levels and high correlations with treatment and anthocyanins were indeed related to stress responses and/or to anthocyanin content (Table 3). Specifically, most gene products of the selected DEGs have been described to participate in the response to one or more types of stresses. Several of the DEGs are involved in the response to biotic stresses, like resistance to bacteria (Lsat_1_v5_gn_1_21441 (Ramírez et al., 2013)), virus (Lsat_1_v5_gn_3_1101 (Tsitsekian et al., 2023)), or fungi (Lsat_1_v5_gn_3_20640 (Wu et al., 2022)), but most have been described to act in abiotic stress responses. In particular, genes related to water deficit and/or the stress-responsive hormone ABA (abscisic acid) stood out, such as Lsat_1_v5_gn_2_116640 (Clément et al., 2011), Lsat_1_v5_gn_2_43400 (Palusa et al., 2007), and Lsat_1_v5_gn_3_20640 (Tahmasebi et al., 2019). Furthermore, not only genes reported to be generally activated under water stress conditions were included in the selection, but also some described as negative regulators, which were in fact downregulated (inhibition of suppressors) in our samples subject to the drought treatment (Tables 2, 3), like Lsat_1_v5_gn_1_50480 (Lee et al., 2022), Lsat_1_v5_gn_2_90361 (Qu et al., 2016), and

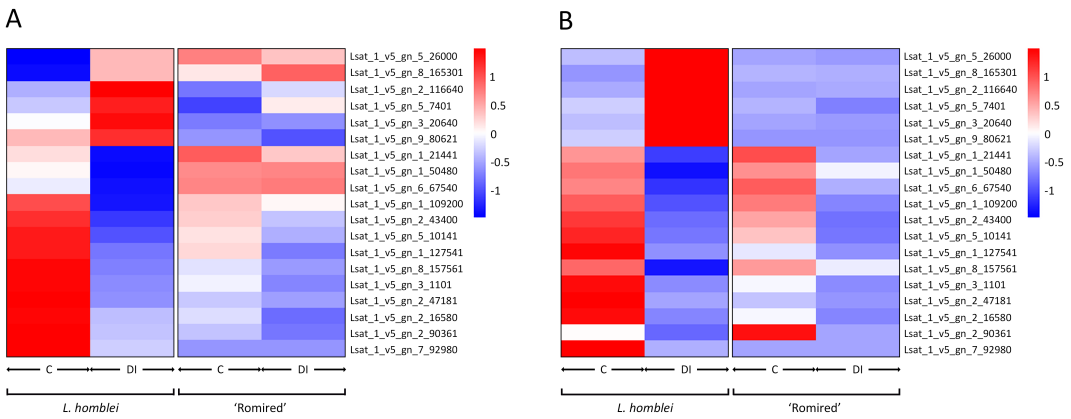


FIGURE 6
Heatmap representation of hierarchical analysis of the mean expression data (n=3) of 19 selected genes in *L. homblei* and 'Romired' under control (C) and deficit irrigation (DI) conditions according to **(A)** RNA-seq and **(B)** real-time qPCR analyses.

Lsat_1_v5_gn_6_67540 (Na and Metzger, 2020). Some of the selected DEGs have been related to other abiotic stresses such as salt (Lsat_1_v5_gn_1_127541 (Bhattarai et al., 2021) and Lsat_1_v5_gn_3_1101 (He et al., 2020)), heat (Lsat_1_v5_gn_2_47181 (Zhang et al., 2022)), and nutrient deficiency (Lsat_1_v5_gn_9_80621 (Zhou et al., 2021)), which makes sense as especially salt and heat stresses often occur simultaneously with drought.

Genes related to flavonoid (e.g., anthocyanins) accumulation were mostly upregulated in our samples (Tables 2 and 3). Two of those genes encode putative transcription factors (TFs): a zinc finger protein (Lsat_1_v5_gn_5_26000 (Zhang et al., 2021)) and a NAC TF (Lsat_1_v5_gn_5_7401 (Wei et al., 2020)), both having been described to induce anthocyanin-related genes. Among the DEGs identified in this work, other two have been previously described to cause the increase of flavonoid or anthocyanin content under abiotic stress. Specifically, Lsat_1_v5_gn_2_15680 was found to play important roles in flavonoid accumulation under drought stress in tea (Li et al., 2020), and Lsat_1_v5_gn_8_165301 was involved in the increase of anthocyanin content when overexpressed in *Arabidopsis thaliana* under ABA-mediated abiotic stress responses (Wang et al., 2018). Another selected gene (Lsat_1_v5_gn_8_157561) is not well characterised but could possibly be related to the anthocyanin content as it contains a MYB domain and, in plants, MYB TFs have been described as one of the major transcriptional regulators of anthocyanin pathway, both activators and repressors (Yan et al., 2021; Cao et al., 2024). This specific gene was downregulated in our samples when exposed to water stress, so it might be a transcriptional repressor of anthocyanin biosynthesis. Lsat_1_v5_gn_7_92980 encodes a ribonuclease III-like protein 3 (RTL3) that cleaves doble-stranded RNA (Comella et al., 2008) and might be also related to anthocyanin content. Proteins with ribonuclease III domains have been described to participate in the regulation of seed coat in soybean and fruit colour in peach through the production of siRNAs (small interfering RNAs) and the increase of transcription levels of genes implied in anthocyanin regulation (Zhu et al., 2012; Jia et al., 2020). Finally, two genes that encode proteins involved in DNA damage repair were also selected (Lsat_1_v5_gn_1_109200 (Fujimori et al., 2014), and Lsat_1_v5_gn_5_10141 (Dangel et al., 2014)). This is not surprising as the generation of ROS is a potential cause of DNA damage under drought stress and a fine-tuned regulation of DNA repair is required to tolerate it (Shim et al., 2018).

3.6 Polymorphisms in the DEGs

In silico search and prediction of polymorphisms were performed to get an overview of the variation in the sequences of the total number of DEGs detected and to explore more deeply the structural variation of the set of 19 selected DEGs.

A total of 235,600 polymorphisms were found in the whole set of DEGs (9,236) in *L. homblei* and 'Romired' compared to the reference genome (both shared and species-exclusive). Most polymorphisms were detected in *L. homblei*, as expected since it is a wild species that is very distant from the cultivated *L. sativa* used as reference. The predominant types of polymorphisms were SNPs (Single Nucleotide Polymorphisms) (89.22%), followed by MNPs (Multiple Nucleotide Polymorphisms) (9.97%), and, in a much smaller extent, by indels (insertions-deletions) (0.81%). The most abundant polymorphism effects were synonymous (63.20%) and missense (27.96%). We also identified intron (6.14%) and splice region (2.23%) variants, as well as others that were present in less than 0.1%, so they are not detailed here.

In the subset of 19 DEGs selected, a total of 404 polymorphisms with 408 predicted effects were identified in *L. homblei* (Table 4), in contrast to the 11 polymorphisms with 12 predicted effects found in those same 19 genes non-differentially expressed in 'Romired' (Supplementary Table S2). Considering only the 19 DEGs in *L. homblei*, the proportions of both polymorphism types and effects were almost the same than those found in the whole set of DEGs, 87.87% of polymorphisms were SNPs, 11.14% were MNPs, and 0.99% were indels. Once again, we found that the predominant predicted effect was synonymous (70.10%), followed by missense (27.44%) type (Table 4). We also identified, though in a reduced number of genes, effects in splice regions (0.98%) and introns (0.49%), disruptive and

TABLE 3 Putative function of the 19 candidate differentially expressed genes (DEGs) in *L. homblei*.

Gene ID	Gene product	Putative function	Reference
Lsat_1_v5_gn_1_21441	Subtilisin-like protease SBT3	Plant immune priming in systemic induced resistance establishment	Ramírez et al., 2013
Lsat_1_v5_gn_1_50480	Haloacid dehalogenase (HAD)-like hydrolase superfamily protein	Repression of ABA-response and ABA-mediated drought tolerance	Lee et al., 2022
Lsat_1_v5_gn_1_109200	DNA damage-repair/toleration protein DRT100	Repair and toleration of UV-B-induced DNA damage	Fujimori et al., 2014
Lsat_1_v5_gn_1_127541	14 kDa proline-rich protein DC2.15	Cell wall modification and organization	Bhattarai et al., 2021
Lsat_1_v5_gn_2_15680	GDSL esterase/lipase	Flavonoid accumulation and lipid reduction under drought stress	Li et al., 2020
Lsat_1_v5_gn_2_43400	Probable pectate lyase 8	Response to stimulus through cell wall modification	Palusa et al., 2007
Lsat_1_v5_gn_2_47181	Protein ECERIFERUM 26	Dehydration tolerance under heat stress	Zhang et al., 2022
Lsat_1_v5_gn_2_90361	Gibberellin-regulated protein 6	ABA-repressible peptide hormone precursor	Qu et al., 2016
Lsat_1_v5_gn_2_116640	Heat shock cognate 70 kDa protein 2	ABA-induced stomatal closure	Clément et al., 2011
Lsat_1_v5_gn_3_1101	Zinc finger protein ZAT1	Putative transcription factor that acts in the response to abiotic and biotic stresses	He et al., 2020; Tsitsekian et al., 2023
Lsat_1_v5_gn_3_20640	PRA1 family protein E	Protein transporter involved in abiotic and biotic stress responses	Tahmasebi et al., 2019; Wu et al., 2022
Lsat_1_v5_gn_5_7401	NAC transcription factor 56	Transcription factor that induces anthocyanin accumulation	Wei et al., 2020
Lsat_1_v5_gn_5_10141	Protein MHF1 homolog	DNA repair and homologous recombination	Dangel et al., 2014
Lsat_1_v5_gn_5_26000	B-box zinc finger protein 21-like	Positive transcriptional regulator of light-induced anthocyanin accumulation	Zhang et al., 2021
Lsat_1_v5_gn_6_67540	Type I inositol polyphosphate 5-phosphatase 2	Putative repressor of water stress response	Na and Metzger, 2020
Lsat_1_v5_gn_7_92980	Ribonuclease III-like protein RTL3	Cleavage of double strand RNA	Comella et al., 2008
Lsat_1_v5_gn_8_157561	Transcription factor MYC/MYB N-terminal domain-containing protein	Putative transcriptional repressor of anthocyanin biosynthesis	Yan et al., 2021
Lsat_1_v5_gn_8_165301	Phospholipase A1 phospholipid-inositol phosphatase PLIP2	ABA-mediated abiotic stress responses and anthocyanin accumulation	Wang et al., 2018
Lsat_1_v5_gn_9_80621	Amino acid permease 6	AA transport under nutrient stresses	Zhou et al., 2021

10.3389/fpls.2024.1494339

10.3389/fpls.2024.1494339

TABLE 4 Predicted effects for the polymorphisms detected in the 19 candidate differentially expressed genes (DEGs) in *L. homblei*.

Gene ID	Conservative in-frame deletion[a]	Disruptive in-frame deletion[a]	Frameshift variant[a]	Intron variant	Missense variant[a]	Splice region variant	Synonymous variant
Lsat_1_v5_gn_1_21441	–	–	–	–	**23**	–	27
Lsat_1_v5_gn_1_50480	–	–	–	–	**3**	–	5
Lsat_1_v5_gn_1_109200	–	–	–	–	**4**	–	27
Lsat_1_v5_gn_1_127541	–	–	–	–	**1**	–	3
Lsat_1_v5_gn_2_15680	–	–	–	–	**3**	–	3
Lsat_1_v5_gn_2_43400	–	–	–	–	**5**	1	31
Lsat_1_v5_gn_2_47181	–	–	–	–	**7**	–	10
Lsat_1_v5_gn_2_90361	–	–	–	–	**2**	–	3
Lsat_1_v5_gn_2_116640	–	**1**	–	2	**4**	–	51
Lsat_1_v5_gn_3_1101	**1**	–	**1**	–	**12**	–	8
Lsat_1_v5_gn_3_20640	–	–	–	–	**4**	–	11
Lsat_1_v5_gn_5_7401	–	–	–	–	**7**	–	16
Lsat_1_v5_gn_5_10141	–	–	–	–	**1**	1	4
Lsat_1_v5_gn_5_26000	–	–	–	–	**8**	–	12
Lsat_1_v5_gn_6_67540	–	–	–	–	**5**	–	11
Lsat_1_v5_gn_7_92980	–	–	–	–	**3**	–	2
Lsat_1_v5_gn_8_157561	–	–	–	–	**3**	2	24
Lsat_1_v5_gn_8_165301	–	**1**	–	–	**13**	–	19
Lsat_1_v5_gn_9_80621	–	–	–	–	**4**	–	19
Percentage	0.25	0.49	0.25	0.49	27.44	0.98	70.10

[a]High and moderate effects are shown in bold.

conservative in-frame deletions (0.49 and 0.25%, respectively), and a frameshift variant (0.25%). The impact of the polymorphisms was frequently low, which makes sense considering that most of them were predicted to have a synonymous effect. However, a polymorphism with high impact was detected. It was a 2-bp insertion that theoretically causes a frameshift mutation in the Lsat_1_v5_gn_3_1101 gene of the wild species which is responsible for the appearance of a premature stop codon. A conservative in-frame deletion was also found in this same gene. According to our results, this gene was downregulated in *L. homblei* whereas in 'Romired' was not differentially expressed, in which showed low expression levels in both C and DI conditions. This gene codes for a zinc-finger protein and appears in the literature as a putative TF that intervenes in the response to abiotic stress (He et al., 2020). A possible effect of one or both polymorphisms might be that the truncated protein acts as a repressor in the wild species under C conditions but stops inhibiting its target(s) as a consequence of its own downregulations under water stress. Other polymorphisms with possible important effects were the putative disruptive in-frame deletions found in Lsat_1_v5_gn_2_116640 and in Lsat_1_v5_gn_8_165301 genes, whose predicted impact was moderate. Lsat_1_v5_gn_2_116640 encodes a 70-kDa heat shock cognate protein. Heat shock proteins (HSPs) were initially described in relation to heat tolerance (Ritossa, 1962), although nowadays they are well known to be expressed in response to a great diversity of environmental stressors besides heat (reviewed in Ul Haq et al. (2019)). According to the RNA-seq analysis, this gene (Lsat_1_v5_gn_2_116640) showed a considerable increase in expression in *L. homblei* in response to drought, whereas in 'Romired' there was not any significant change, with the values under both C and DI similar to those in C plants of *L. homblei*. Therefore, the disruptive in-frame deletion in this gene could be inducing the activation of this HSP when *L. homblei* plants are subject to drought stress. Lsat_1_v5_gn_8_165301 encodes a Phospholipase A1 phospholipid-inositol phosphatase 2 (PLIP2) that has been described to be involved in the accumulation of anthocyanins under ABA-mediated abiotic stress responses (Wang et al., 2018). This gene also exhibited a highly significant upregulation in *L. homblei* and no change of expression in 'Romired'. In this case, its expression levels in 'Romired' under C and DI conditions were similar to those of *L. homblei* under DI. Thus, the disruptive in-frame deletion found in *L. homblei* sequence might be causing the gene to be activated only under stress in the wild plants.

4 Discussion

4.1 Identification and analysis of DEGs under drought stress conditions

The number of exclusive up- and downregulated genes was more than twice in *L. homblei* and 1.3 times higher in 'Romired', respectively. In general, genes related to regulation within biological process (e.g., response to water) and molecular function categories were more abundant and more intensively upregulated in the wild species whereas genes responsible for cellular components were more commonly and significantly upregulated in the cultivated species. In the case of the downregulated DEGs, the most represented terms in both species were those related to catalytic activities. In general terms, basal and growth-related processes were deactivated in both species, which probably contributes to redirect resources to guarantee plant survival.

Interestingly, activation of responses seemed to be species specific, and it looks like the CWR was triggering more mechanisms of response to drought stress as the number of upregulated genes was clearly larger in *L. homblei* under DI. In contrast, the higher number of downregulated DEGs common to both species, many of them implied in basal processes, could be due to the deactivation of basal metabolism processes to designate more resources to water deficit tolerance, previously described in different plant species subject to water stress (Shao et al., 2009). The results from the GO enrichment analysis are in agreement with those found in other studies that assessed different stresses in lettuce, in which response to stimulus, biological regulation, metabolic processes, binding and catalytic activities, as well as membrane components, were the most represented terms (Wang et al., 2017; Zhou et al., 2023). Interestingly, in a transcriptomic analysis carried out to identify genes involved in lettuce anthocyanin accumulation, the most represented GO terms were the same (Zhang et al., 2016).

Both anthocyanin contents and the number of upregulated genes were clearly larger in *L. homblei* under DI. Therefore, results point to a relationship between gene expression profiles (for some DEGs) and changes in the accumulation of these antioxidant compounds. In addition, not only the most abundant anthocyanins showing the biggest change in quantity in response to water stress but also the minor anthocyanins only identified under DI (peonidin 3-*O*-glucoside and cyanidin 3-(6''-acetylglucoside) in *L. homblei* and 'Romired', respectively) could play a role in the response to drought. That is more plausible in the case of *L. homblei* where the differences in anthocyanin content between C and DI conditions were significant (Medina-Lozano et al., 2024).

Interestingly, to activate the anthocyanin biosynthesis route in *L. homblei* seems to be enough to upregulate the isoforms of the gene controlling the first step of the specific pathway branch (i.e., *CHS*), even when the preceding genes from the general phenylpropanoid pathway (e.g., 4-coumarate-CoA ligase (*4CL*)) could be downregulated or not significantly differentially expressed, as they are involved in the biosynthesis of many other phenylpropanoids apart from anthocyanins. Something similar has been shown in a previous study on the expression of those genes and the anthocyanin content of poplar leaves (Tian et al., 2021).

All these differences between both *Lactuca* spp. might reflect a lager plasticity of the wild species to adapt to environmental changes. The great genetic diversity of wild species allows them to counteract the effects of different stresses more effectively (Jordanovska et al., 2020), whereas the cultivated species could have lost these mechanisms through domestication. In fact, the common DEGs to both accessions which show an opposite sense in the change of expression could consist of genes that have acquired a different mode of action as *L. homblei* belongs to the lettuce tertiary gene pool, the most genetically distant from *L. sativa.* Alternatively, they could be artefacts, either methodological (e.g., library preparation) or statistical or even both.

4.2 Validation, putative function, and polymorphisms of candidate DEGs

The fact that the genes with the strongest change of expression and correlation with anthocyanin content and drought were differentially expressed only in *L. homblei* might reveal, once again, the existence of tolerance mechanisms in the wild species that are not present in the cultivated one. This is in agreement with the wild species showing the highest increase (and the only resulting statistically significant) of anthocyanins in a previous study on drought stress with the same accessions, among others (Medina-Lozano et al., 2024).

The reliability of the results from the RNA-seq analysis was confirmed as the candidate DEGs were validated by real-time qPCR, being all differentially expressed only in the wild species, in which the expression profiles obtained with the two techniques coincided. Besides, most gene products of the selected DEGs have been described to participate in the response to one or more types of stresses which makes sense as some stresses often occur simultaneously.

The fact that the genes with a high change in the level of expression identified in this study resulted to be related to both the stress response and the anthocyanin content could indicate that these compounds are playing an important role for plants to cope with the drought conditions, as was also proposed before in purple-stem *Brassica napus* L. (Chen et al., 2022b).

Talking about the polymorphisms found in all DEGs and in the candidate genes, our results are in agreement with studies carried out in other crops that also used transcriptomic data, in which the number of SNPs was also much higher than the number of indels, and the most abundant polymorphism effects were synonymous and missense variants too (Iquebal et al., 2017; Muñoz-Espinoza et al., 2020). The impact of the polymorphisms identified in the candidate DEGs frequently resulted low as most of them were predicted to have a synonymous effect. However, a few of them showed a high or moderate predicted impact what could be a reflection of the drastic changes in the gene expression profiles (either activation or inhibition) in the wild species when subject to

drought stress. Among the rest of the polymorphisms found, the missense variants could have an impact on the function of the resultant protein, though this one has been predicted to be moderate.

In silico tools are truly useful for obtaining information of functional and structural variants on the transcriptome and their possible correlation with phenotypic changes (Yazar and Özbek, 2021). However, further experimental approaches like functional analyses are essential to verify in the future the polymorphism effect found in putative candidate genes involved in the anthocyanin accumulation and the response to drought stress.

5 Conclusion

Mechanisms of response to drought stress related to anthocyanins were triggered in the wild species *L. homblei* but not in the cultivated lettuce variety 'Romired'. The involvement of the proposed candidate genes in the increase of anthocyanin content and the response to drought stress in the wild species is supported by their large and significant changes in the expression levels when the plants were subjected to water deprivation and by their high correlation with anthocyanin content. Furthermore, the activation of the anthocyanin biosynthesis route was mainly achieved by significantly upregulating the genes controlling the first step of the specific branch (flavonoid pathway), again exclusively in the wild species.

All the candidate genes have been reported before to be involved in the response to biotic or abiotic stresses in other species (but not in lettuce), what demonstrate that plants have developed interconnected and interacting routes to deploy integrated responses to combinations of concurrent stresses.

This wild species has become a potential donor of drought tolerance genes to the cultivated lettuce that foreseeably will make the crop more resilient and sustainable, while containing more beneficial compounds (i.e., anthocyanins) for human health.

Declaration of Generative AI and AI-assisted technologies in the writing process'

During the preparation of this work the authors have not used any AI-assisted technology in order to generate the manuscript.

Data availability statement

The datasets presented in this study can be found in the European Nucleotide Archive (ENA) at https://www.ebi.ac.uk/ena/browser/home accession number: PRJEB75159.

Author contributions

IML: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JG: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Writing – review & editing. AD: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the projects PID2022-138484OR-I00 from the Spanish Ministry of Science and Innovation and State Research Agency (AEI) and LMP148_21 from the Government of Aragón; and by the Operational Programme FEDER Aragón 2023-2025 and 2020-2022, and the European Social Fund from the European Union (A12-23R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética"). IML was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Spanish State Research Agency (AEI).

Acknowledgments

We thank J. A. Aranjuelo for technical support and D. L. Goodchild for reviewing the English language. We gratefully acknowledge the Centre for Genetic Resources (CGN, Wageningen, Netherlands) for supplying the seeds used here.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1494339/full#supplementary-material

SUPPLEMENTARY TABLE 1
Primer sequences, amplicon length, and annealing temperature of the reference gene (*TRXL3-3*) and the 19 differentially expressed genes (DEGs) selected to be validated by qPCR.

SUPPLEMENTARY TABLE 2
Predicted effects for the polymorphisms detected in the 19 candidate differentially expressed genes (DEGs) in 'Romired'.

SUPPLEMENTARY FIGURE 1
Expression data obtained by qPCR (relative expression) and by RNA-seq (CPM, counts per million) of 19 selected genes in the lettuce commercial variety 'Romired' under control (C) and deficit irrigation (DI) conditions. Bars represent standard error of the mean (n=3). Transformation $1/(1+x)^2$ was applied to achieve normal distribution in qPCR data in the following cases: Lsat_1_v5_gn_1_21441, Lsat_1_v5_gn_2_43400, Lsat_1_v5_gn_2_90361, and Lsat_1_v5_gn_6_67540. Wilcoxon test was used with non-normally distributed qPCR data of Lsat_1_v5_gn_1_127541, Lsat_1_v5_gn_2_15680, Lsat_1_v5_gn_2_47181, Lsat_1_v5_gn_3_1101, Lsat_1_v5_gn_3_20640, and Lsat_1_v5_gn_5_10141.

SUPPLEMENTARY FIGURE 2
Heatmap representation of hierarchical analysis of the expression data of 19 selected differentially expressed genes (DEGs) in *L. homblei* and 'Romired' under control (C) and deficit irrigation (DI) conditions according to (A) RNA-seq and (B) real-time qPCR analyses. Numbers 1-3 and 4-6 represent the biological replicates under C and DI, respectively.

References

Becker, C., Klaering, H. P., Kroh, L. W., and Krumbein, A. (2014). Cool-cultivated red leaf lettuce accumulates cyanidin-3-*O*-(6″-*O*- malonyl)-glucoside and caffeoylmalic acid. *Food Chem.* 146, 404–411. doi: 10.1016/j.foodchem.2013.09.061

Bhattarai, S., Fu, Y. B., Coulman, B., Tanino, K., Karunakaran, C., and Biligetu, B. (2021). Transcriptomic analysis of differentially expressed genes in leaves and roots of two alfalfa (*Medicago sativa* L.) cultivars with different salt tolerance. *BMC Plant Biol.* 21, 446. doi: 10.1186/s12870-021-03201-4

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cao, Y., Mei, Y., Zhang, R., Zhong, Z., Yang, X., Xu, C., et al. (2024). Transcriptional regulation of flavonol biosynthesis in plants. *Hortic. Res.* 11, uhae043. doi: 10.1093/hr/uhae043

Castellarin, S. D., Pfeiffer, A., Sivilotti, P., Degan, M., Peterlunger, E., and Di Gaspero, G. (2007). Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ.* 30, 1381–1399. doi: 10.1111/j.1365-3040.2007.01716.x

Chen, L., Xu, M., Liu, C., Hao, J., Fan, S., and Han, Y. (2022a). LsMYB15 regulates bolting in leaf lettuce (*Lactuca sativa* L.) under high-temperature stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.921021

Chen, W., Miao, Y., Ayyaz, A., Hannan, F., Huang, Q., Ulhassan, Z., et al. (2022b). Purple stem *Brassica napus* exhibits higher photosynthetic efficiency, antioxidant potential and anthocyanin biosynthesis related genes expression against drought stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.936696

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Clément, M., Leonhardt, N., Droillard, M.-J., Reiter, I., Montillet, J.-L., Genty, B., et al. (2011). The cytosolic/nuclear HSC70 and HSP90 molecular chaperones are important for stomatal closure and modulate abscisic acid-dependent physiological responses in *Arabidopsis*. *Plant Physiol.* 156, 1481–1492. doi: 10.1104/pp.111.174425

Comella, P., Pontvianne, F., Lahmy, S., Vignols, F., Barbezier, N., DeBures, A., et al. (2008). Characterization of a ribonuclease III-like protein required for cleavage of the pre-rRNA in the 3′ETS in *Arabidopsis*. *Nucleic Acids Res.* 36, 1163–1175. doi: 10.1093/nar/gkm1130

Dangel, N. J., Knoll, A., and Puchta, H. (2014). MHF1 plays Fanconi anaemia complementation group M protein (FANCM)-dependent and FANCM-independent roles in DNA repair and homologous recombination in plants. *Plant J.* 78, 822–833. doi: 10.1111/tpj.12507

Eriksen, R. L., Knepper, C., Cahn, M. D., and Mou, B. (2016). Screening of lettuce germplasm for agronomic traits under low water conditions. *HortScience* 51, 669–679. doi: 10.21273/hortsci.51.6.669

Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., Dinger, M. E., Hellemans, J., et al. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.* 7, 1559. doi: 10.1038/s41598-017-01617-3

FAO. (2021). *The impact of disasters and crises on agriculture and food security: 2021* (Rome, Italy). doi: 10.4060/cb3673en

FAOSTAT. (2021). *Statistics of the Food and Agriculture Organization of the United Nations*. Available online at: http://www.fao.org/faostat/en/data/QC (Accessed January 10, 2024).

Faria, D. (2017). *GOEnrichment* (GitHub Repos). Available online at: https://github.com/DanFaria/GOEnrichment (Accessed February 10, 2024).

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., and Basra, S. M. A. (2009). "Plant drought stress: effects, mechanisms and management," in *Sustainable Agriculture*. Eds. E. Lichtfouse, M. Navarrete, P. Debaeke, S. Véronique and C. Alberola (Springer, Dordrecht), 153–188. doi: 10.1007/978-90-481-2666-8_12

Fujimori, N., Suzuki, N., Nakajima, Y., and Suzuki, S. (2014). Plant DNA-damage repair/toleration 100 protein repairs UV-B-induced DNA damage. *DNA Repair.* 21, 171–176. doi: 10.1016/j.dnarep.2014.05.009

Garcia, C., and Blesso, C. N. (2021). Antioxidant properties of anthocyanins and their mechanism of action in atherosclerosis. *Free Radic. Biol. Med.* 172, 152–166. doi: 10.1016/j.freeradbiomed.2021.05.040

Garrison, E. P. (2015).vcflib. In: *GitHub repository* (GitHub). Available online at: https://github.com/ekg/vcflib (Accessed February 10, 2024).

Garrison, E. P., and Marth, G. T. (2012). *Haplotype-based variant detection from short-read sequencing* (Ithaca, United States: Cornell University). Available at: http://arxiv.org/abs/1207.3907 (Accessed February 09, 2024).

He, F., Niu, M. X., Feng, C. H., Li, H. G., Su, Y., Su, W. L., et al. (2020). *PeSTZ1* confers salt stress tolerance by scavenging the accumulation of ROS through regulating the expression of *PeZAT12* and *PeAPX2* in *Populus*. *Tree Physiol.* 40, 1292–1311. doi: 10.1093/treephys/tpaa050

Horvath, S., and Dong, J. (2008). Geometric interpretation of gene coexpression network analysis. *PloS Comput. Biol.* 4, e1000117. doi: 10.1371/journal.pcbi.1000117

Hugueney, P., Provenzano, S., Verriès, C., Ferrandino, A., Meudec, E., Batelli, G., et al. (2009). A novel cation-dependent *O*-methyltransferase involved in anthocyanin methylation in grapevine. *Plant Physiol.* 150, 2057–2070. doi: 10.1104/pp.109.140376

Ino, I., and Yamaguchi, M. A. (1993). Acetyl-coenzyme A: Anthocyanidin 3-glucoside acetyltransferase from flowers of *Zinnia elegans*. *Phytochemistry* 33, 1415–1417. doi: 10.1016/0031-9422(93)85101-V

IPCC (2021). "Summary for policymakers," in *Climate Change 2021. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. Vol. 2 . Eds. V. Masson-Delmotte, P. Zhai and A. Pirani (Cambridge, United Kingdom: Cambridge University Press).

Iquebal, M. A., Soren, K. R., Gangwar, P., Shanmugavadivel, P. S., Aravind, K., Singla, D., et al. (2017). Discovery of putative herbicide resistance genes and its regulatory network in chickpea using transcriptome sequencing. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00958

Jia, J., Ji, R., Li, Z., Yu, Y., Nakano, M., Long, Y., et al. (2020). Soybean DICER-LIKE2 regulates seed coat color via production of primary 22-nucleotide small interfering RNAs from long inverted repeats. *Plant Cell* 32, 3662–3673. doi: 10.1105/tpc.20.00562

Jordanovska, S., Jovovic, Z., and Andjelkovic, V. (2020). "Potential of wild species in the scenario of climate change," in *Rediscovery of Genetic and Genomic Resources for Future Food Security*. Eds. R. K. Salgotra and S. M. Zargar (Springer Nature, Singapore), 263–302. doi: 10.1007/978-981-15-0156-2_10

Ju, Y., Yang, B., He, S., Tu, T., Min, Z., Fang, Y., et al. (2019). Anthocyanin accumulation and biosynthesis are modulated by regulated deficit irrigation in Cabernet Sauvignon (*Vitis vinifera* L.) grapes and wines. *Plant Physiol. Biochem.* 135, 469–479. doi: 10.1016/j.plaphy.2018.11.013

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Koyama, R., Yoshimoto, A., Ishibashi, M., Itoh, H., and Uno, Y. (2021). Enzymatic activities and gene transcript levels associated with the augmentation of antioxidant constituents during drought stress in lettuce. *Horticulturae* 7, 444. doi: 10.3390/horticulturae7110444

Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Lee, S., Choi, E., Kim, T., Hwang, J., and Lee, J.-H. (2022). AtHAD1, A haloacid dehalogenase-like phosphatase, is involved in repressing the ABA response. *Biochem. Biophys. Res. Commun.* 587, 119–125. doi: 10.1016/j.bbrc.2021.11.095

Li, M., Liu, J., Zhou, Y., Zhou, S., Zhang, S., Tong, H., et al. (2020). Transcriptome and metabolome profiling unveiled mechanisms of tea (*Camellia sinensis*) quality improvement by moderate drought on pre-harvest shoots. *Phytochemistry* 180, 112515. doi: 10.1016/j.phytochem.2020.112515

Liao, Y., Smyth, G. K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Llorach, R., Martínez-Sánchez, A., Tomás-Barberán, F. A., Gil, M. I., and Ferreres, F. (2008). Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chem.* 108, 1028–1038. doi: 10.1016/j.foodchem.2007.11.032

Medina-Lozano, I., Arnedo, M. S., Grimplet, J., and Díaz, A. (2023). Selection of novel reference genes by RNA-seq and their evaluation for normalising real-time qPCR expression data of anthocyanin-related genes in lettuce and wild relatives. *Int. J. Mol. Sci.* 24, 3052. doi: 10.3390/ijms24033052

Medina-Lozano, I., Bertolín, J. R., and Díaz, A. (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* 359, 129864. doi: 10.1016/j.foodchem.2021.129864

Medina-Lozano, I., Bertolín, J. R., and Díaz, A. (2024). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1369658

Moreno-Escamilla, J. O., Jiménez-Hernández, F. E., Alvarez-Parrilla, E., de la Rosa, L. A., Martínez-Ruiz, N. D. R., González-Fernández, R., et al. (2020). Effect of Elicitation on Polyphenol and Carotenoid Metabolism in Butterhead Lettuce (*Lactuca sativa* var. capitata). *ACS Omega* 5, 11535–11546. doi: 10.1021/acsomega.0c00680

Mou, B. (2005). Genetic variation of beta-carotene and lutein contents in lettuce. *J. Am. Soc Hortic. Sci.* 130, 870–876. doi: 10.21273/jashs.130.6.870

Muñoz-Espinoza, C., Di Genova, A., Sánchez, A., Correa, J., Espinoza, A., Meneses, C., et al. (2020). Identification of SNPs and InDels associated with berry size in table grapes integrating genetic and transcriptomic approaches. *BMC Plant Biol.* 20, 1–21. doi: 10.1186/s12870-020-02564-4

Na, J.-K., and Metzger, J. D. (2020). A putative tomato inositol polyphosphate 5-phosphatase, Le5PT1, is involved in plant growth and abiotic stress responses. *3 Biotech.* 10, 28. doi: 10.1007/s13205-019-2023-y

Naing, A. H., and Kim, C. K. (2021). Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. *Physiol. Plant* 172, 1711–1723. doi: 10.1111/ppl.13373

Paim, B. T., Crizel, R. L., Tatiane, S. J., Rodrigues, V. R., Rombaldi, C. V., and Galli, V. (2020). Mild drought stress has potential to improve lettuce yield and quality. *Sci. Hortic. (Amsterdam)* 272, 109578. doi: 10.1016/j.scienta.2020.109578

Palusa, S. G., Golovkin, M., Shin, S. B., Richardson, D. N., and Reddy, A. S. N. (2007). Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in *Arabidopsis*. *New Phytol.* 174, 537–550. doi: 10.1111/j.1469-8137.2007.02033.x

Park, S., Shi, A., and Mou, B. (2020). Genome-wide identification and expression analysis of the CBF/DREB1 gene family in lettuce. *Sci. Rep.* 10, 5733. doi: 10.1038/s41598-020-62458-1

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45. doi: 10.1093/nar/29.9.e45

PGR (Plant Genetic Resources) Lettuce. *The lettuce gene pool*. Available online at: https://www.pgrportal.nl/en/lettuce-genetic-resources-portal.htm (Accessed January 27, 2024).

Qu, J., Kang, S. G., Hah, C., and Jang, J. C. (2016). Molecular and cellular characterization of GA-Stimulated Transcripts GASA4 and GASA6 in *Arabidopsis thaliana*. *Plant Sci.* 246, 1–10. doi: 10.1016/j.plantsci.2016.01.009

Quezada-Martinez, D., Addo Nyarko, C. P., Schiessl, S. V., and Mason, A. S. (2021). Using wild relatives and related species to build climate resilience in *Brassica* crops. *Theor. Appl. Genet.* 134, 1711–1728. doi: 10.1007/s00122-021-03793-3

Ramírez, V., López, A., Mauch-Mani, B., Gil, M. J., and Vera, P. (2013). An extracellular subtilase switch for immune priming in *Arabidopsis*. *PloS Pathog.* 9, e1003445. doi: 10.1371/journal.ppat.1003445

Reddy, A. R., Chaitanya, K. V., and Vivekanandan, M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J. Plant Physiol.* 161, 1189–1202. doi: 10.1016/j.jplph.2004.01.013

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8, 14953. doi: 10.1038/ncomms14953

Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia* 18, 571–573. doi: 10.1007/BF02172188

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rugienius, R., Bendokas, V., Siksnianas, T., Stanys, V., Sasnauskas, A., and Kazanaviciute, V. (2021). Characteristics of *Fragaria vesca* Yield Parameters and Anthocyanin Accumulation under Water Deficit Stress. *Plants* 10, 557. doi: 10.3390/plants10030557

Shao, H. B., Chu, L. Y., Jaleel, C. A., Manivannan, P., Panneerselvam, R., and Shao, M. A. (2009). Understanding water deficit stress-induced changes in the basic metabolism of higher plants-biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. *Crit. Rev. Biotechnol.* 29, 131–151. doi: 10.1080/07388550902869792

Shim, J. S., Oh, N., Chung, P. J., Kim, Y. S., Choi, Y. D., and Kim, J. K. (2018). Overexpression of OsNAC14 improves drought tolerance in rice. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00310

Su, W., Tao, R., Liu, W., Yu, C., Yue, Z., He, S., et al. (2020). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* 18, 479–490. doi: 10.1111/pbi.13213

Tahmasebi, A., Ashrafi-Dehkordi, E., Shahriari, A. G., Mazloomi, S. M., and Ebrahimie, E. (2019). Integrative meta-analysis of transcriptomic responses to abiotic stress in cotton. *Prog. Biophys. Mol. Biol.* 146, 112–122. doi: 10.1016/j.pbiomolbio.2019.02.005

Tange, O. (2011). *GNU Parallel: The Command-Line Power Tool* Vol. 36 (Frederiksberg, Denmark: Login USENIX Mag), 42–47. Available at: https://www.gnu.org/software/parallel/. (Accessed February 09, 2024)

The Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* 50, W345–W351. doi: 10.1093/nar/gkac247

Tian, Y., Li, Q., Rao, S., Wang, A., Zhang, H., Wang, L., et al. (2021). Metabolic profiling and gene expression analysis provides insights into flavonoid and anthocyanin metabolism in poplar. *Tree Physiol.* 41, 1046–1064. doi: 10.1093/treephys/tpaa152

Tsitsekian, D., Daras, G., Templalexis, D., Avgeri, F., Lotos, L., Orfanidou, C. G., et al. (2023). A subset of highly responsive transcription factors upon tomato infection by pepino mosaic virus. *Plant Biol.* 25, 529–540. doi: 10.1111/plb.13515

Tsormpatsidis, E., Henbest, R. G. C., Davis, F. J., Battey, N. H., Hadley, P., and Wagstaffe, A. (2008). UV irradiance as a major influence on growth, development and secondary products of commercial importance in Lollo Rosso lettuce "Revolution" grown under polyethylene films. *Environ. Exp. Bot.* 63, 232–239. doi: 10.1016/j.envexpbot.2007.12.002

Ul Haq, S., Khan, A., Ali, M., Khattak, A. M., Gai, W.-X., Zhang, H.-X., et al. (2019). Heat shock proteins: dynamic biomolecules to counter plant biotic and abiotic stresses. *Int. J. Mol. Sci.* 20, 5321. doi: 10.3390/ijms20215321

Wada, K. C., Inagaki, N., Sakai, H., Yamashita, H., Nakai, Y., Fujimoto, Z., et al. (2022). Genetic effects of Red Lettuce Leaf genes on red coloration in leaf lettuce under artificial lighting conditions. *Plant Environ. Interact.* 3, 179–192. doi: 10.1002/pei3.10089

Wang, Y., Chen, R., Hao, Y., Liu, H., Song, S., and Sun, G. (2017). Transcriptome analysis reveals differentially expressed genes (DEGs) related to lettuce (*Lactuca sativa*) treated by TiO2/ZnO nanoparticles. *Plant Growth Regul.* 83, 13–25. doi: 10.1007/s10725-017-0280-5

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Wang, K., Guo, Q., Froehlich, J. E., Hersh, H. L., Zienkiewicz, A., Howe, G. A., et al. (2018). Two abscisic acid-responsive plastid lipase genes involved in jasmonic acid biosynthesis in *Arabidopsis thaliana*. *Plant Cell* 30, 1006–1022. doi: 10.1105/tpc.18.00250

Wei, Z., Hu, K., Zhao, D. L., Tang, J., Huang, Z. Q., Jin, P., et al. (2020). MYB44 competitively inhibits the formation of the MYB340-bHLH2-NAC56 complex to regulate anthocyanin biosynthesis in purple-fleshed sweet potato. *BMC Plant Biol.* 20, 258. doi: 10.1186/s12870-020-02451-y

Wickham, H. (2009). *ggplot2 - Elegant Graphics for Data Analysis* (New York, NY, USA: Springer). doi: 10.1007/978-3-319-24277-4

Wu, N., Li, W. J., Chen, C., Zhao, Y. P., and Hou, Y. X. (2022). Analysis of the PRA1 genes in cotton identifies the role of GhPRA1.B1-1A in *Verticillium dahliae* resistance. *Genes (Basel)* 13, 765. doi: 10.3390/genes13050765

Xiong, T., Zhang, S., Kang, Z., Zhang, T., and Li, S. (2021). Dose-Dependent Physiological and Transcriptomic Responses of Lettuce (*Lactuca sativa* L.) to Copper Oxide Nanoparticles-Insights into the Phytotoxicity Mechanisms. *Int. J. Mol. Sci.* 22, 3688. doi: 10.3390/ijms22073688

Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M., et al. (2021). MYB-mediated regulation of anthocyanin biosynthesis. *Int. J. Mol. Sci.* 22, 3103. doi: 10.3390/ijms22063103

Yazar, M., and Özbek, P. (2021). *In silico* tools and approaches for the prediction of functional and structural effects of single-nucleotide polymorphisms on proteins: an expert review. *Omi. A J. Integr. Biol.* 25, 23–37. doi: 10.1089/omi.2020.0141

Zeljković, S. C., Štefelová, N., Hron, K., Doležalová, I., and Tarkowski, P. (2023). Preharvest abiotic stress affects the nutritional value of lettuce. *Agronomy* 13, 398. doi: 10.3390/agronomy13020398

Zhang, F., Rosental, L., Ji, B., Brotman, Y., and Dai, M. (2024). Metabolite-mediated adaptation of crops to drought and the acquisition of tolerance. *Plant J.* 118, 626–644. doi: 10.1111/tpj.16634

Zhang, H., Wan, Z., Liu, J., Hu, X., Ren, L., Feng, S., et al. (2022). DsCER26 affects the leaf dehydration tolerance of rice by altering cuticular wax alkane production without affecting the grain fatty acid content. *ACS Agric. Sci. Technol.* 2, 813–822. doi: 10.1021/acsagscitech.2c00141

Zhang, Y., Xu, S., Cheng, Y., Peng, Z., and Han, J. (2018). Transcriptome profiling of anthocyanin-related genes reveals effects of light intensity on anthocyanin biosynthesis in red leaf lettuce. *PeerJ* 13, e4607. doi: 10.7717/peerj.4607

Zhang, Y. Z., Xu, S. Z., Cheng, Y. W., Ya, H. Y., and Han, J. M. (2016). Transcriptome analysis and anthocyanin-related genes in red leaf lettuce. *Genet. Mol. Res.* 15, gmr.15017023. doi: 10.4238/gmr.15017023

Zhang, B., Zhu, Z. Z., Qu, D., Wang, B. C., Hao, N. N., Yang, Y. Z., et al. (2021). MdBBX21, a B-box protein, positively regulates light-induced anthocyanin accumulation in apple peel. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.774446

Zhou, H., Yu, L., Liu, S., Zhu, A., Yang, Y., Chen, C., et al. (2023). Transcriptome comparison analyses in UV-B induced AsA accumulation of *Lactuca sativa* L. *BMC Genomics* 24, 61. doi: 10.1186/s12864-023-09133-7

Zhou, T., Yue, C., Huang, J., Cui, J., Liu, Y., Wang, W., et al. (2021). Genome-wide identification of the amino acid permease genes and molecular characterization of their transcriptional responses to various nutrient stresses in allotetraploid rapeseed. *BMC Plant Biol.* 21, 151. doi: 10.1186/s12870-021-03043-0

Zhu, H., Xia, R., Zhao, B., An, Y., Dardick, C. D., Callahan, A. M., et al. (2012). Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. *BMC Plant Biol.* 12, 149. doi: 10.1186/1471-2229-12-149

CHAPTER 4.2. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1494339/full#supplementary-material.

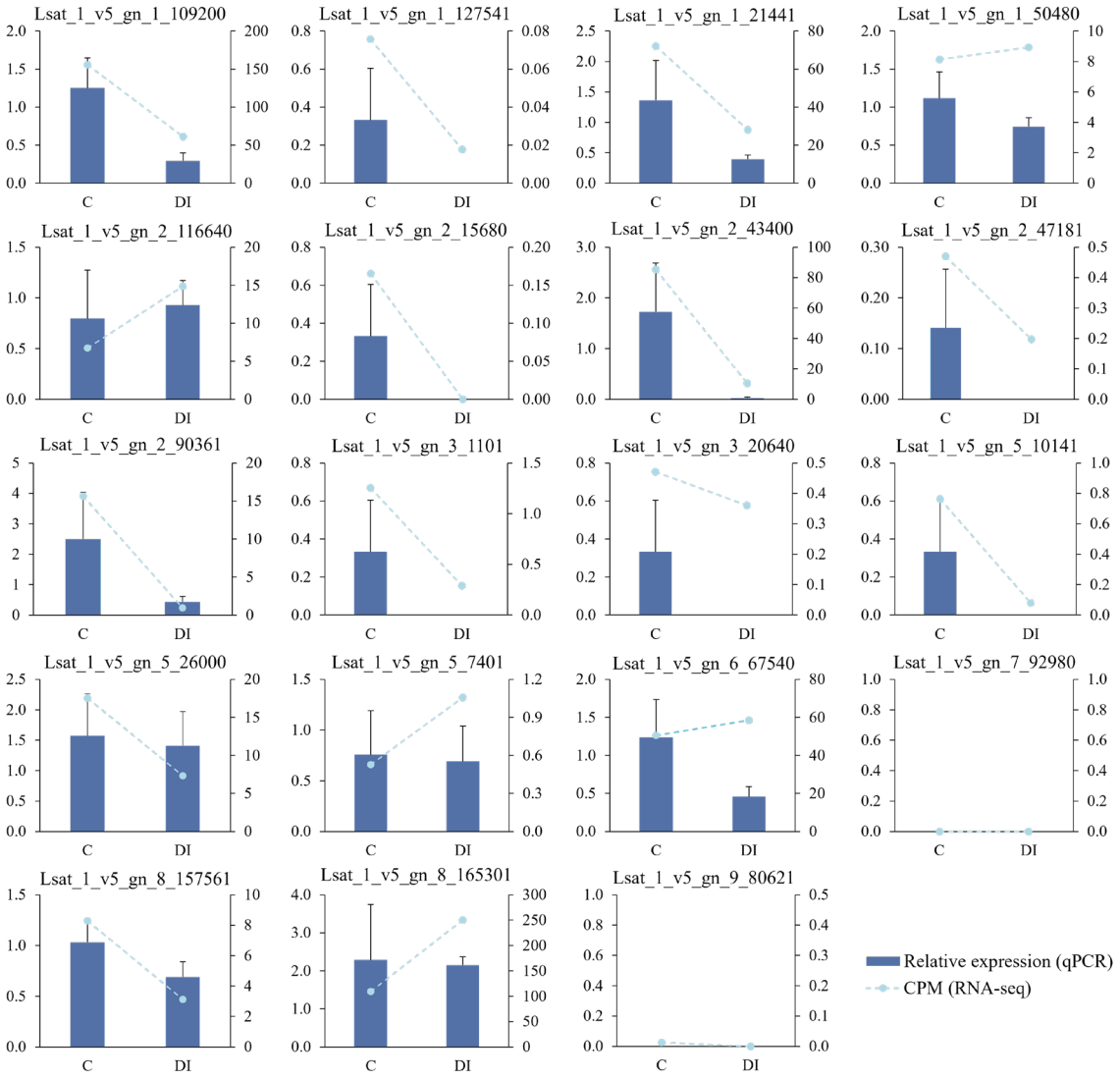


Figure S1. Expression data obtained by qPCR (relative expression) and by RNA-seq (CPM, counts per million) of 19 selected genes in the lettuce commercial variety 'Romired' under control (C) and deficit irrigation (DI) conditions. Bars represent standard error of the mean (n = 3). Transformation $1/(1+x)^2$ was applied to achieve normal distribution in qPCR data in the following cases: Lsat_1_v5_gn_1_21441, Lsat_1_v5_gn_2_43400, Lsat_1_v5_gn_2_90361, and Lsat_1_v5_gn_6_67540. Wilcoxon test was used with non-normally distributed qPCR data of Lsat_1_v5_gn_1_127541, Lsat_1_v5_gn_2_15680, Lsat_1_v5_gn_2_47181, Lsat_1_v5_gn_3_1101, Lsat_1_v5_gn_3_20640, and Lsat_1_v5_gn_5_10141.

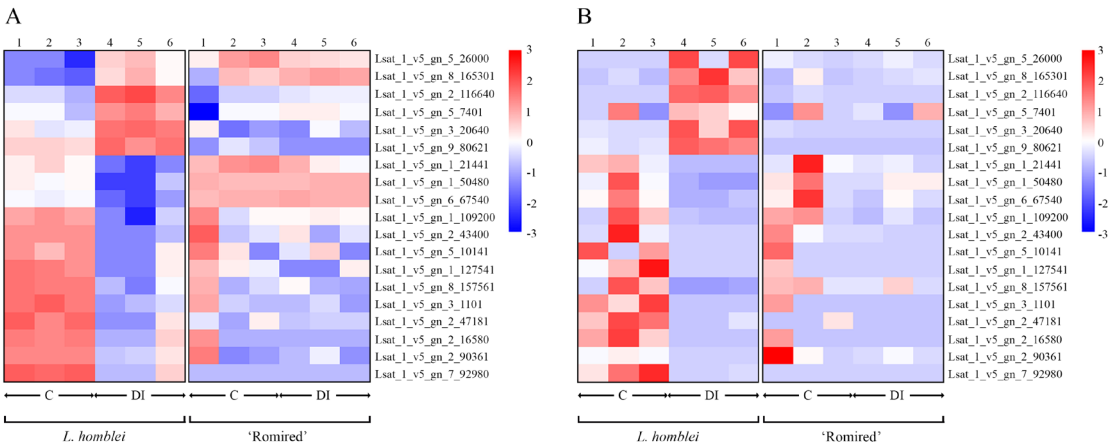


Figure S2. Heatmap representation of hierarchical analysis of the expression data of 19 selected differentially expressed genes (DEGs) in *L. homblei* and 'Romired' under control (C) and deficit irrigation (DI) conditions according to (A) RNA-seq and (B) real-time qPCR analyses. Numbers 1-3 and 4-6 represent the biological replicates under C and DI, respectively.

Table S1. Primer sequences, amplicon length, and annealing temperature of the reference gene (*TRXL3-3*) and the 19 differentially expressed genes (DEGs) selected to be validated by qPCR.

Name	Primer sequence (5' - 3')	Amplicon length (bp)	Annealing temperature (°C)
Lsat_1_v5_gn_1_109200	F-GCGTGTATAACTTCTCTCCCT R-TTTCCCTAACTTTCCGATG	96	60
Lsat_1_v5_gn_1_127541	F-AGCCACATAAACATCACCACCA R-CGACAAGGTCCCCAAGGAG	150	66
Lsat_1_v5_gn_1_21441	F-AAAGGTGAATGTGAAGAAGG R-AGCGTTAAGTTTGGGTGA	113	57
Lsat_1_v5_gn_1_50480	F-AGCGAAGAAGAAAAAGAAGAAA R-CCAAGAGGAATCACAAGCAG	110	60
Lsat_1_v5_gn_2_116640	F-AACGTCTTATCGGTGATGCT R-GCGAATTGTTTCTCTTCTCCT	205	63
Lsat_1_v5_gn_2_15680	F-TTCGAGAATGTATGATGTGGG R-GAGGGTTGAATAAATTGGCTG	146	60
Lsat_1_v5_gn_2_43400	F-GGTGGTTAAGGAATCGGTGT R-CTTTGCTCTTTCTGCCATTG	163	60
Lsat_1_v5_gn_2_47181	F-CGAGTCAGTAGATGATGACAA R-GCAGAAAAAACATCTCCAA	165	57
Lsat_1_v5_gn_2_90361	F-CGCTGTCTTCCTCATAGCTC R-GCTTGTCAAACTTCCAGGTC	103	63
Lsat_1_v5_gn_3_1101	F-CATTTAGTCACCTTACCCATTC R-CTCCCTTAATCCGTACATCA	159	54
Lsat_1_v5_gn_3_20640	F-CCATGTTTTTGATGATCGTGT R-CCTGATTCCGGCGTGT	142	52
Lsat_1_v5_gn_5_10141	F-ACGGAAGCGATGAGGAGA R-GGAGATGCAAACGATAACAGG	146	60
Lsat_1_v5_gn_5_26000	F-GCGATTCATAAAGTCAACC R-AGCATCTCAATCAAGTACTCC	278	59
Lsat_1_v5_gn_5_7401	F-GGACGAAAACGAATTGG R-AACCCAATCATCCAACCTCAA	119	56
Lsat_1_v5_gn_6_67540	F-AGGTTAATGATCGGGACTTGG R-CGGGGCACTATAGCTTTTAGG	252	56
Lsat_1_v5_gn_7_92980	F-CCATTGCACTCCACTGGTT R-CCACCTCCCATGTAGCATC	118	63
Lsat_1_v5_gn_8_157561	F-CTCCTCTCAAAACTCCCATCA R-CACCATCCTTCACATCTTCCT	169	60
Lsat_1_v5_gn_8_165301	F-TGCATGTAGCGTTGATGATT R-GGCTTGATTTGAGGTATGGA	153	59
Lsat_1_v5_gn_9_80621	F-TTTTTAGATTGGTGTGGA R-AACCCTAAGAAGCTATTGA	88	54
TRXL3-3	F-TGGTGTCGTGTTTGTGCAGAG R-TTGGGTTGTTTCTGGGCATT	111	62

Table S2. Predicted effects for the polymorphisms detected in the 19 candidate differentially expressed genes (DEGs) in 'Romired'.

Gene ID	Intron variant	Missense variant[a]	Splice region variant	Synonymous variant
Lsat_1_v5_gn_1_21441	-	-	-	-
Lsat_1_v5_gn_1_50480	-	-	-	-
Lsat_1_v5_gn_1_109200	-	-	-	-
Lsat_1_v5_gn_1_127541	-	-	-	-
Lsat_1_v5_gn_2_15680	-	-	-	1
Lsat_1_v5_gn_2_43400	-	**1**	-	-
Lsat_1_v5_gn_2_47181	-	-	-	-
Lsat_1_v5_gn_2_90361	-	-	-	-
Lsat_1_v5_gn_2_116640	-	-	-	-
Lsat_1_v5_gn_3_1101	-	-	-	-
Lsat_1_v5_gn_3_20640	-	-	-	-
Lsat_1_v5_gn_5_7401	-	-	-	-
Lsat_1_v5_gn_5_10141	-	-	-	-
Lsat_1_v5_gn_5_26000	-	-	-	1
Lsat_1_v5_gn_6_67540	1	-	-	-
Lsat_1_v5_gn_7_92980	-	-	-	-
Lsat_1_v5_gn_8_157561	-	-	-	-
Lsat_1_v5_gn_8_165301	1	**1**	1	5
Lsat_1_v5_gn_9_80621	-	-	-	-
Percentage	16.67	16.67	8.33	58.33

[a]Moderate effects are shown in bold

GENERAL DISCUSSION

GENERAL DISCUSSION

Despite the economic importance of lettuce, the growing interest in healthier diets and the concerns about the devastating effects of climate change on agriculture, nutritional quality and tolerance to abiotic stresses have been scarcely considered as breeding objectives for this crop. Since the beginnings of lettuce breeding, the main aims have been to introduce resistance to the predominant diseases and to increase the yield; the latter actually became a major goal in most crop breeding programmes after the Green Revolution. This, together with the elimination of compounds conferring unpopular flavours to maximize customer satisfaction, resulted in a nutrient "wash" in cultivated lettuce. To help covering the gap in both nutritional quality and tolerances to abiotic stresses in lettuce breeding, this thesis aims, on one hand, at characterizing the nutritional value of a diverse lettuce-related germplasm to select a variety rich in vitamin C, antioxidant compound with benefits for both plant defence and human health, as well as at dissecting, to some extent, the genetic control of vitamin C accumulation in lettuce. On the other hand, it aims at studying the changes in vitamin C and anthocyanin contents in response to drought stress in different *Lactuca* spp. using different analytical approaches.

1. Quantification of vitamin C and anthocyanins in lettuce germplasm

Evaluation of the content in the phytochemicals of interest should be the first step in a breeding programme directed to improve the nutritional value and/or the health-promoting properties of an agri-food crop. That characterization is essential for the selection of the plant material to get a broad representation of phytochemicals. So, we characterized 30 lettuce-related accessions in terms of vitamin C and anthocyanin content in this thesis, including green and red commercial lettuce varieties, green and semi-red traditional lettuce varieties and some wild *Lactuca* spp. [Chapter 1 (Medina-Lozano et al., 2021)]. In addition, both compounds were characterized in two different tissues, leaves and stems, of two CWR with an early bushy growth (*L. dregeana* and *L. squarrosa*). To the best of our knowledge, this was the first time that both compounds were assessed in such a diverse lettuce-related germplasm. Some studies have quantified them in a smaller number of accessions, being all of them commercial varieties, as reviewed in Kim et al. (2016). In the work carried out by van Treuren et al. (2018), AA was quantified in a great number of accessions of primitive and modern lettuce varieties

and wild relatives, including the three most commonly used in research, *L. serriola, L. saligna* and *L. virosa*, as well as some underused species. Unlike in most of the previous works, in our study, not only AA but also the total vitamin C (AA and DHAA) was assessed and, as far as we know, some accessions, both cultivated lettuce varieties and wild species, were characterized for the first time for vitamin C and anthocyanin content. The interest in bioactive compounds of lettuce landraces and traditional varieties is growing in the last few years as it can be noted in more recent researches (Martínez-Ispizua et al., 2022; Zeljković et al., 2023).

Another important aspect to consider is the analytical method used to evaluate the phytochemical content. Vitamin C and anthocyanin quantification has been conducted using different approaches in a great number of crops. They are labile compounds affected by different environmental factors like intense light, elevated temperatures, high pH, etc. (Chalker-Scott, 1999; Lee and Kader, 2000). That is why we examined their stability. In the case of anthocyanins, they remained stable over a 24-h period, whereas vitamin C degradation started 4 h after its extraction [Chapter 1 (Medina-Lozano et al., 2021) and Annex 4 (Medina-Lozano et al., 2020)]. Consequently, we needed a robust method that ensured an efficient extraction and quantification of vitamin C in its two forms (AA and DHAA) and that enabled the analysis of the maximum number of samples in the shortest possible time to assess accurately and quickly all our samples. Therefore, an optimized UPLC-UV protocol for vitamin C quantification was developed in this thesis [Annex 4 (Medina-Lozano et al., 2020)].

In Chapter 1 (Medina-Lozano et al., 2021) we found that the average content of vitamin C or TAA was higher in the CWR, followed by the traditional varieties and, finally, by the commercial lettuces. The opposite ranking was obtained in the case of mean anthocyanin content, that is, the richest accessions were the commercial varieties, followed distantly by the traditional ones and then by the CWR. This is a clear reflection of the historical breeding objectives in lettuce. Anthocyanins, apart from having health-promoting properties, confer the leaves a characteristic colour, which increases the market value of the red varieties and hence, have been considered an agronomic trait of interest, appealing for both farmers and consumers and, obviously, also for breeders. Selection for this phenotype has resulted in an indirect increase in anthocyanins, which has led to obtaining commercial varieties richer in these compounds, as we observed here and also

agrees with the results reported by Casals Missio et al. (2018). On the contrary, nutritional quality has been generally put aside in lettuce breeding, and this is consistent with the lower vitamin C content observed in commercial varieties. The highest amounts were found in the assessed CWR and then in the traditional varieties, which have escaped totally and partially from strong selection pressures, respectively, in accordance with results obtained by van Treuren et al. (2018).

When considering exclusively the cultivated lettuces, differences in the amount of vitamin C present depending on the leaf colour were obtained. Vitamin C content was higher in the green-leaf varieties than in the red-leaf ones. This could be another proof of how selection for an agronomic character (colour) has a negative impact on nutrient accumulation. The explanation for these differences might lie in the metabolic costs associated with phytochemical biosynthesis in plants. Red varieties, whose anthocyanin content has been enhanced, could allocate more resources to the biosynthesis of these compounds at the expense of vitamin C, whereas green lettuces, lacking anthocyanins, can direct those resources towards vitamin C biosynthesis. This is supported by the negative correlation that we observed between the amount of both compounds [Chapter 1 (Medina-Lozano et al., 2021)]. García-Macías et al. (2007) also suggested a reduction in some plant processes due to anthocyanin biosynthesis because of a direct competition for assimilated carbon.

In Chapter 1 (Medina-Lozano et al., 2021), a classification of the lettuce-related accessions was established based on their content in both compounds, vitamin C and anthocyanins. Three groups were distinguished in both cases, which were those with high, medium and low content. The richest and intermediate groups in vitamin C consisted mainly of traditional lettuce varieties and wild relatives, reflecting their value from a nutritional perspective. On the contrary, the richest group in anthocyanins was composed exclusively of red commercial varieties, which is further evidence of the enhancement in anthocyanin content as a consequence of red colour selection in lettuce breeding programmes. This classification of lettuce-related germplasm could offer valuable information not only for farmers and consumers, but also for breeders.

An opposite trend between both compounds was also observed when comparing the phytochemical content in the tissues of the two bushy-growth CWR. In both species, vitamin C content was higher in the leaves, while anthocyanins were more abundant in

the stems. Maybe this could also be explained by a trade-off in resource allocation. Vitamin C, that participates in photosynthesis, is present in higher amounts than anthocyanins in the main photosynthetic organ (i.e., leaves), while more resources are allocated to accumulate anthocyanins in stems, where vitamin C is likely not so necessary.

All results considered, Chapter 1 (Medina-Lozano et al., 2021) establishes the starting point of the two following research lines of the thesis. On one hand, the highest concentration of vitamin C was detected in a plant of 'Lechuga del Pirineo'. That is why this traditional variety was chosen to be used with two different aims. Firstly, to obtain a genetically homogeneous variety with high vitamin C content and, secondly, to identify marker-trait associations and putative candidate genes responsible for vitamin C accumulation in lettuce. On the other hand, a set of lettuce-related accessions was selected, initially, to study the response to drought stress in terms of vitamin C and anthocyanin content, and subsequently, to analyse the expression changes in genes related to those antioxidant compounds whose quantity increased under the stress.

2. Selection of a lettuce variety rich in vitamin C, genetic diversity analysis and identification of marker-vitamin C content associations in cultivated lettuce germplasm

'Lechuga del Pirineo' is a Cos-type lettuce from the Aragonese Pyrenees (Spain), known for its good agronomic and organoleptic characteristics (Carravedo et al., 2011) and, according to the results obtained in Chapter 1 (Medina-Lozano et al., 2021), also with good nutritional quality (high vitamin C content). These attributes, together with the growing interest of consumers in local products and the advantages inherent to traditional varieties, that is, their great genetic diversity in comparison with the commercial varieties and the absence of sexual barriers and linkage drag in contrast to CWR, make 'Lechuga del Pirineo' an ideal candidate for a lettuce breeding programme aimed at its biofortification. Thus, breeding populations of 'Lechuga del Pirineo' were generated during this thesis to obtain a genetically homogeneous variety with good agronomic and nutritional characteristics [Chapter 2 (Medina-Lozano et al., 2024b)]. The first population (S0) consisted of seeds supplied by the BGHZ (Spain), so likely coming from different plants, whereas the two following populations (S1 and S2) were obtained by self-fecundation of the richest plant in vitamin C from S0 and S1, respectively. Vitamin C

content and genetic variability revealed increased homogeneity in the successive generations, as expected. Indeed, genetic homogeneity for the assessed set of markers was already reached at S1 (obviously, S2 was also genetically homogeneous). Although average TAA content was lower in S2, it still fell in the range of the "high TAA" group according to the classification established in Chapter 1 (Medina-Lozano et al., 2021) and the spread of the data was reduced considerably, what is desirable for a potential new variety (higher homogeneity). Besides, AA, the most biologically active form of the vitamin, increased highly significantly with the two self-pollination rounds.

Due to the importance of vitamin C in the nutritional quality of plant-based food, we conducted a GWAS to find genetic associations with this trait as well as putative candidate genes related to the accumulation of the vitamin. S0 was the selected population of 'Lechuga del Pirineo' for the GWAS as it was the only one that harboured genetic variability. Even so, the assessed markers resulted to be highly monomorphic in S0, what makes sense taking into account the predominantly autogamous nature of lettuce. Therefore and following Hamazaki et al. (2020) recommendations, a diversity panel was included in the GWAS to increase its power. Adding variability to a target population allows to increase the heterozygosity needed in this kind of studies and provides a better genome coverage with more polymorphic markers. The diversity panel consisted of 21 cultivated lettuces, including commercial and traditional varieties. Genetic diversity analyses showed that the panel was composed of three different groups, one consisting exclusively of red commercial lettuces, other formed of only green traditional varieties and a third one more diverse that was the largest and included commercial and traditional varieties, mainly green and semi-red. Different analyses indicated that the first mentioned group was the most distinguishable, whereas the second and the third seemed to be closer. Indeed, results showed that traditional varieties harbour greater variability, as some of them clustered separately from the others even when they shared certain characteristics, such as the crop type (Cos lettuces), the colour (green) and the small area of origin (Aragonese Pyrenees).

Including the S0 population of 'Lechuga del Pirineo' and the diversity panel of cultivated lettuces in the GWAS, a set of SNPs significantly associated with DHAA was identified [Chapter 2 (Medina-Lozano et al., 2024b)]. Traits related to nutritional value and/or health-promoting compounds have been barely explored by GWAS in lettuce,

except for one work assessing primary metabolite content (Zhang et al., 2020) and others targeting anthocyanins although as an agronomic character (colour) and not as phytochemicals with health beneficial properties (Kwon et al., 2013; Wei et al., 2021). Therefore, to the best of our knowledge, the associations between markers and vitamin C content identified in this thesis are the first described in lettuce and they are among the first described in other crops when referring specifically to DHAA. As far as we know, marker-trait associations with DHAA have only been detected in a previous study conducted in tomato (Sauvage et al., 2014). AA has been prioritized in crop breeding because it is the most biologically active form of vitamin C and it is easier to measure using chromatographic methods, but DHAA importance should not be underestimated. DHAA also shows biological activity itself and it is easily recyclable to AA in the human body, serving as a vitamin C reservoir, as proposed in Chapter 1 (Medina-Lozano et al., 2021).

To gain more insight, LD was studied in-depth in the genomic region of chromosome 2 where the significant associations were identified. Interestingly, among all the SNPs present in such region, high LD values of the lead SNP (the highest significantly associated marker to DHAA content) were precisely obtained with only other significantly associated SNPs. Results suggested that there could be more than one locus causing the phenotypic variation as the highest LD values were not necessarily found for either the most significantly associated or the physically closest SNPs (the lack of order is a possible indication of the existence of multiple signals). Vitamin C content can be considered a notably complex trait (Venkatesh and Park, 2014), so the presence of more than one gene and/or causative variant makes sense. Besides, the genes that lead to the high DHAA content in lettuce may not be those containing the significantly associated polymorphisms, but others in high LD with them but not covered with any polymorphic marker. Therefore, every gene found in the region of interest was explored, and actually several candidate genes, for example, those encoding pectinesterases, F-box proteins or long noncoding RNA, have previously been described to be related to vitamin C content in other crops like potato (Berdugo-Cely et al., 2023), tomato (Ruggieri et al., 2015) and wheat (Zhou et al., 2015). More research is needed to unravel their roles in vitamin C metabolism in lettuce. However, this study helps to obtain a more comprehensive understanding of the genetics underlying this important trait, which could result useful in a breeding programme aimed at lettuce biofortification.

3. Study of vitamin C and anthocyanin contents in response to drought stress in *Lactuca* spp.

3.1. Vitamin C and anthocyanin quantification in control and drought stress conditions

In addition to addressing the lack of knowledge on the nutritional quality of lettuce, it is of the utmost importance to explore the crop tolerance to abiotic stresses, especially drought, which is one of the most concerning environmental threats for agriculture in the present scenario of climate change and it has been barely tackled in lettuce breeding to date. Consequently, drought stress impact on vitamin C and anthocyanin contents was assessed in lettuce-related germplasm [Chapter 3 (Medina-Lozano et al., 2024a)], as these compounds show antioxidant activity potentially beneficial for plant protection against different stresses. Plant material for the study was selected based on the classification established in Chapter 1 (Medina-Lozano et al., 2021) to cover a wide range of variability in the quantity of both compounds, from accessions belonging to the poorest group to those within the richest. Thus, five commercial lettuce varieties (red and green) and two wild relatives (including one with early bushy growth, *L. dregeana*, to assess different tissues, leaves and stems) were included in the analysis. The study was conducted in two consecutive years and both control and drought stress conditions were evaluated [Chapter 3 (Medina-Lozano et al., 2024a)]. The obtained data followed consistent tendencies across the two years and treatments. In the same way, results under control conditions were in line with those obtained in Chapter 1 (Medina-Lozano et al., 2021). That is, vitamin C was again more abundant in the tested lettuce wild relatives and in the leaf tissue of *L. dregeana*, whereas anthocyanin content was higher in the commercial varieties (considering only the red ones as green varieties do not contain anthocyanins) and in the stems of *L. dregeana*. When subject to drought stress, an opposite trend for both compounds was observed in the leaf samples: vitamin C decreased while anthocyanins increased in all the tested accessions. In the case of the stem tissue of *L. dregeana*, vitamin C accumulation decreased, as in the leaves. Quite the opposite, the effect of drought on anthocyanin amount in the stems was the contrary than in the leaves, being the only case in which these compounds decreased with the stress. These results suggest that anthocyanins (unlike vitamin C) may play a crucial role in lettuce tolerance to drought. Moreover, the decrease of anthocyanin content in the stem and its increase in the leaves of *L. dregeana* under water deficit may indicate a quick

response to the stress via compound translocation towards where they are more needed, although more research on anthocyanin subcellular transport is necessary. Those results were confirmed in the three biological repeats of *L. dregeana*. Remarkably, despite the increase in anthocyanin content in all the assessed accessions, it was only significant in one CWR (*L. homblei*) both years. This reflects the higher responsiveness to water stress of this CWR. Thus, *L. homblei* could be a valuable source of drought tolerance and so, a good plant material for breeding programmes aimed at enhancing lettuce resilience as well as the content in health-promoting compounds even if techniques such as bridge crosses could be needed as it belongs to GP3. At the same time, the potential new varieties would be more sustainable as they could have less water necessities thanks to the improved tolerance to water deficit.

Regarding individual compounds, the anthocyanins detected in the drought stress experiment reported in Chapter 3 (Medina-Lozano et al., 2024a) were the same previously identified [Chapter 1 (Medina-Lozano et al., 2021)]. Likewise, cyanidin 3-*O*-(6'-*O*-malonylglucoside) was the most abundant and the only one present in all the accessions, tissues and treatments. However, some differences were observed between both studies concerning the minor anthocyanins. On one hand, cyanidin 3-(6''-acetylglucoside) had been exclusively detected in the stem samples of the CWR *L. squarrosa*, whereas in the most recent study, apart from in the stems (not in the leaves) of *L. dregeana*, it was also detected in two cultivated lettuces, but only under drought conditions in one of them ('Romired'). On the other hand, peonidin 3-*O*-glucoside had only been found in *L. sativa* in the first study, but in the most recent one, it was also identified in a wild relative (*L. homblei*), exclusively under drought conditions, like cyanidin 3-(6''-acetylglucoside) in 'Romired', as just mentioned. These minor anthocyanins could have been biosynthesized in very specific circumstances and may also be key in the acquisition of drought tolerance in lettuce, though a more thorough investigation is needed to elucidate it.

The drop in vitamin C content under water deficit, also reported previously in lettuce (Zeljković et al., 2023), reinforces the hypothesis of resource allocation suggested before. As vitamin C does not seem to participate in lettuce response against drought, its biosynthesis might be reduced to prioritize anthocyanin biosynthesis (in red varieties), which might actually help to combat the stress. In green lettuces, lacking anthocyanins, vitamin C content also decreased under drought stress. In these cases, resources might be

allocated towards other protection processes, such as the biosynthesis of other flavonoids which have been reported to increase in some green varieties subject to water deficit (Zeljković et al., 2023). That is, processes that are not implied in drought tolerance and so, in plant survival, like it could be vitamin C biosynthesis, might be suppressed when plants need to face the stress.

3.2. Transcriptomic analysis

Getting a better understanding of the genetic basis behind anthocyanin changes in *Lactuca* spp. response to drought stress is essential to achieve the long-term goal of enhancing both its tolerance and its health-promoting properties. The importance of this approach is even higher considering the scarcity of transcriptomic studies in lettuce aimed at analysing metabolite-related genes involved in the response to water deficit. For this purpose, an RNA-seq analysis focused on the response of anthocyanin-related genes to drought was conducted in a commercial lettuce variety and a wild relative [Chapter 4.2 (Medina-Lozano et al., 2025)]. Plant material was selected based on results of Chapter 3 (Medina-Lozano et al., 2024a), in which all tested accessions experienced an increased in their anthocyanin contents in response to water stress. One representative member of the cultivated species and another of the wild relatives were included in the study. 'Romired' was the chosen commercial variety because it showed the lowest variance among replicates, crucial in RNA-seq analyses to ensure the identification of higher number of DEGs. The CWR was *L. homblei* as it was the only accession that showed a significant or very significant anthocyanin increase under stress among all the assessed accessions. Another reason that contributed to their selection was that one of the minor anthocyanins appeared exclusively under deficit irrigation conditions in each accession (cyanidin 3-(6''-acetylglucoside) in 'Romired' and peonidin 3-*O*-glucoside in *L. homblei*).

Similarly to what happened with the increased anthocyanin content (only significant in the wild species), *L. homblei* was the accession that triggered the highest anthocyanin-related response to drought stress at transcriptomic level. This was reflected in different ways. First of all, the number of upregulated DEGs was higher in *L. homblei*, whereas the number of downregulated DEGs was higher in 'Romired'. GO terms related to regulation processes, response to water included, were also more abundantly and significantly enriched in *L. homblei*. Likewise, the number of significantly upregulated

DEGs involved in the anthocyanin pathway was considerably higher in the wild relative than in the cultivated species. Finally, a set of candidate genes that showed the biggest change of expression and the highest and most significant correlation with both anthocyanin content and drought treatment, were selected. Interestingly, all of them resulted to be differentially expressed only in *L. homblei* and not in 'Romired'. Moreover, we found that some of these candidate genes have functions related to both the response to different stresses, drought included, and flavonoid accumulation, mainly anthocyanins. These results reveal a possible relationship between the differential expression of some genes and the anthocyanin accumulation in a lettuce wild relative under drought stress, as it has already been reported in other crops like grapevine (Castellarin et al., 2007) and canola (Chen et al., 2022). All this reinforces the idea that *L. homblei* is a valuable resource for lettuce breeding.

Furthermore, the validation of the expression of the selected candidate DEGs by other technique, in this case, qPCR, was undertaken because, although RNA-seq is very precise, some artefacts can be found among the obtained data (Everaert et al., 2017). A crucial step in the analysis of gene expression via qPCR is the selection of suitable RGs to ensure an accurate data normalisation and so, a reliable quantification of gene expression. RG suitability can be specific for each experimental context. For a drought stress assay, novel candidate RGs had previously been proposed and tested in lettuce (Borowski et al., 2014). However, their selection was based on genes reported to have a stable expression in other crops. In this thesis new candidate RGs, selected based on our RNA-seq data [Chapter 4.1 (Medina-Lozano et al., 2023) and Annex 5 (Medina-Lozano et al., 2022)], have been proposed. This approach has already been proved to be successful in other crops like apple (Zhou et al., 2017) and soybean (Yim et al., 2015), but as far as we know, this is the first time that has been adopted in lettuce. Among those with low variation in expression levels according to RNA-seq data, six candidate RGs were preselected and their stability was evaluated using different statistical approaches. Then, we provided a list of the most appropriated genes not only for a drought stress experiment but also for studies comparing different leaf colours (green and red) and tissues (leaf and stem) in lettuce-related germplasm. Using the most stable RG for the drought stress assay, gene expression of the selected candidate DEGs was indeed validated since the same expression patterns were obtained using both technologies, RNA-seq and qPCR, confirming the reliability of the RNA-seq analysis.

Finally, polymorphisms within these candidate genes were searched *in silico*. A substantially higher number of variants were found in *L. homblei* than in 'Romired', what makes sense as the reference genome has been obtained in the cultivated species (*L. sativa* 'Salinas' v8 (Reyes-Chin-Wo et al., 2017)) to which 'Romired' belongs. Those variants predicted to have high or moderate impact (115 among the candidate DEGs) could be responsible for the differential expression in the wild relative.

A functional validation of candidate genes potentially responsible for the accumulation of anthocyanins in the wild species *L. homblei* when subject to drought is needed. Physiological studies to establish the level of drought stress that ensures a higher anthocyanin production and tolerance acquisition without compromising lettuce yield should be also conducted. Nevertheless, this thesis establishes some molecular basis that could assist in developing new bred lettuce varieties.

CONCLUSIONS

CONCLUSIONS

Objective 1. Quantification of vitamin C and anthocyanins in different lettuce plant material including commercial and traditional varieties as well as wild relatives (*Lactuca* spp.).

1. The highest mean vitamin C content was found in the lettuce wild relatives and the lowest in the red lettuce commercial varieties. In contrast, the highest mean anthocyanin content was detected in the red commercial varieties and the lowest in the wild relatives. The traditional varieties occupied an intermediate position in both cases.
2. AA was the predominant form of vitamin C in most of the accessions assayed. Among the three anthocyanins detected, cyanidin 3-*O*-(6'-*O*-malonylglucoside) was the most abundant and the only one present in all the accessions and tissues. The other two anthocyanins, cyanidin 3-(6''-acetylglucoside) and peonidin 3-*O*-glucoside, were found rarely and in much less quantity.
3. The richest plant in vitamin C belonged to 'Lechuga del Pirineo', the traditional variety chosen to generate the breeding populations. Genetic homogeneity was already reached in S1 and AA content increased highly significantly in S2.

Objective 2. Selection of the variety with the highest vitamin C content and genetic characterization along with GWAS of vitamin C content in cultivated germplasm (*L. sativa*).

4. Population structure and genetic diversity and relationships of a panel of 21 lettuce varieties consistently showed two groups very well differentiated formed by red commercial and green traditional varieties, respectively, in contrast to another integrated by all types of lettuce varieties with high admixture.
5. H_O values were generally low as expected in a predominantly autogamous species like lettuce. Even though, traditional varieties harboured a considerable diversity even when all came from a small geographical area.
6. A GWAS of vitamin C content conducted on the 'Lechuga del Pirineo' S0 population and 21 cultivated lettuce varieties allowed us to identify significant genetic associations with DHAA content within a 5.1 Mb region of chromosome 2.

7. Only SNPs significantly associated with DHAA content showed high LD values with the lead SNP. LD patterns suggested that more than one locus could be responsible for the variation in DHAA content.
8. Some candidate genes previously reported to be related to vitamin C content in other crops, like those coding for a pectinesterase/pectinesterase inhibitor, several F-box proteins and long noncoding RNAs, were identified in the associated region to DHAA content.

Objective 3. Evaluation of drought stress impact on vitamin C and anthocyanin contents in lettuce varieties and wild relatives.

9. Drought stress caused a decrease in vitamin C in both the commercial lettuce varieties and the wild relatives and in the two tissues (leaves and stems) studied.
10. Drought stress resulted in an accumulation of anthocyanins in leaves of all lettuce commercial varieties and wild relatives analysed though the increase was significant only in one of the wild species (*L. homblei*).
11. Changes in anthocyanin content due to water stress were the opposite in the two assayed tissues of the wild species *L. dregeana*; a decrease in the stem and an increase in the leaves. A translocation of these compounds towards the tissue where they were likely more needed (i.e., leaves) could be happening in response to drought.
12. The anthocyanin-related transcriptomic response to drought stress was clearly triggered in the wild relative *L. homblei* but not in the cultivated variety 'Romired', in line with the significant increase of anthocyanin content only in the wild species.
13. The activation of anthocyanin biosynthesis was mainly through the upregulation of genes controlling the first step of the specific branch (flavonoid pathway).
14. Some proposed candidate genes for the anthocyanin-mediated response to drought have previously been reported to act in mechanisms of resistance and/or tolerance to biotic and abiotic stresses, respectively. This supports the idea of plants deploying interacting routes to activate integrated responses to the combination of different stresses.
15. *L. homblei* could be a potential donor of drought tolerance alleles to the cultivated lettuce to obtain varieties with enhanced resilience, richer in bioactive compounds, and potentially more sustainable by having less water requirements.

CONCLUSIONES

Objetivo 1. Cuantificación de vitamina C y antocianinas en diferentes accesiones de lechuga, incluyendo variedades comerciales y tradicionales, así como parientes silvestres (*Lactuca* spp.).

1. El mayor contenido en vitamina C se encontró en los parientes silvestres de lechuga y el más bajo en las variedades comerciales rojas. Por el contrario, la mayor concentración de antocianinas se detectó en las variedades comerciales rojas y la más baja en los parientes silvestres. Las variedades tradicionales ocuparon una posición intermedia en ambos casos.
2. AA fue la forma de vitamina C predominante en la mayoría de las accesiones estudiadas. De las tres antocianinas detectadas, la cianidina 3-*O*-(6'-*O*-malonilglucósido) fue la más abundante y la única presente en todas las accesiones y tejidos. Las otras dos antocianinas, cianidina 3-(6''-acetilglucósido) y peonidina 3-*O*-glucósido, se encontraron rara vez y en mucha menor cantidad.
3. La planta más rica en vitamina C pertenecía a la variedad tradicional 'Lechuga del Pirineo', que fue la accesión escogida para generar las poblaciones de mejora. La homogeneidad genética se alcanzó en S1 y el contenido en AA aumentó de forma altamente significativa en S2.

Objetivo 2. Selección de la variedad con el contenido más alto en vitamina C y caracterización genética junto con GWAS del contenido en vitamina C en germoplasma cultivado (*L. sativa*).

4. La estructura poblacional y la diversidad y relaciones genéticas de un panel de 21 variedades de lechuga mostraron consistentemente dos grupos muy bien diferenciados formados por variedades comerciales rojas y tradicionales verdes, respectivamente, en contraste con otro compuesto por todos los tipos de variedades de lechuga con alto grado de mezcla.
5. Los valores de H_O fueron por lo general bajos, como es de esperar en una especie predominantemente autógama como la lechuga. Aun así, las variedades tradicionales albergaron una diversidad considerable, especialmente teniendo en cuenta que provienen todas de una región geográfica pequeña.

6. Un GWAS del contenido en vitamina C llevado a cabo en la población S0 de 'Lechuga del Pirineo' y en 21 variedades de lechuga cultivada nos permitió identificar asociaciones genéticas significativas con el contenido en DHAA en una región de 5.1 Mb del cromosoma 2.
7. Solo SNPs asociados significativamente con el contenido en DHAA mostraron valores de LD altos con el SNP centinela. Los patrones de LD sugieren que podría haber más de un locus responsable de la variación en el contenido de DHAA.
8. Se identificaron en la región asociada al contenido en DHAA algunos genes candidatos previamente descritos en relación con el contenido en vitamina C en otros cultivos, como los que codifican una pectinesterasa/inhibidor de pectinesterasa, varias proteínas F-box y ARNs largos no codificantes.

Objetivo 3. Evaluación del impacto del déficit hídrico en el contenido en vitamina C y antocianinas en variedades de lechuga y parientes silvestres.

9. El estrés por sequía provocó una reducción del contenido en vitamina C tanto en las variedades comerciales de lechuga como en los parientes silvestres y en los dos tejidos (hojas y tallos) estudiados.
10. El estrés por sequía resultó en una acumulación de antocianinas en las hojas de todas las variedades comerciales de lechuga y de los parientes silvestres estudiados, aunque el incremento solo fue significativo en una de las especies silvestres (*L. homblei*).
11. Los cambios en el contenido en antocianinas debidos al estrés hídrico fueron opuestos en los tejidos analizados de la especie silvestre *L. dregeana*, disminuyendo en el tallo y aumentando en las hojas. Se podría estar produciendo una traslocación de estos compuestos hacia el tejido donde probablemente son más necesarios (es decir, las hojas) en respuesta a la sequía.
12. La respuesta a nivel transcriptómico al estrés por sequía relacionada con las antocianinas se desencadenó claramente en el pariente silvestre *L. homblei*, pero no en la variedad de lechuga cultivada 'Romired', en consonancia con el aumento significativo del contenido en antocianinas solo en la especie silvestre.
13. La biosíntesis de antocianinas se activó principalmente mediante el incremento de la transcripción de los genes que controlan el primer paso de la rama específica (ruta de los flavonoides).

14. Se han propuesto genes candidatos para la respuesta a sequía mediada por antocianinas, cuya participación en mecanismos de resistencia y/o tolerancia a estreses bióticos y abióticos, respectivamente, ha sido descrita previamente. Esto apoya la idea de que las plantas despliegan rutas que interaccionan para activar respuestas integradas a la combinación de diferentes tipos de estrés.
15. *L. homblei* podría ser un potencial donante de alelos de tolerancia a sequía para la lechuga cultivada, con el fin de obtener variedades con mayor resiliencia, más ricas en compuestos bioactivos y potencialmente más sostenibles al tener menos requerimientos de agua.

REFERENCES

REFERENCES

Adhikari, ND, Simko, I, Mou, B (2019). Phenomic and Physiological Analysis of Salinity Effects on Lettuce. *Sensors* 19, 4814. doi: 10.3390/s19214814

Andersen, ØM, Jordheim, M (2010). Anthocyanins, in *Encyclopedia of Life Sciences (ELS)*, (Chichester: John Wiley & Sons, Ltd). doi: 10.1002/9780470015902.a0001909.pub2

Apel, K, Hirt, H (2004). Reactive Oxygen Species: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701

Baek, GY, Kim, MH, Kim, CH, Choi, EG, Jin, BO, Son, JE, et al. (2013). The effect of LED light combination on the anthocyanin expression of lettuce. *IFAC Proc. Vol.* 46, 120–123. doi: 10.3182/20130327-3-jp-3017.00029

Becker, C, Klaering, HP, Kroh, LW, Krumbein, A (2014). Cool-cultivated red leaf lettuce accumulates cyanidin-3-*O*-(6″-*O*-malonyl)-glucoside and caffeoylmalic acid. *Food Chem.* 146, 404–411. doi: 10.1016/j.foodchem.2013.09.061

Berdugo-Cely, JA, Céron-Lasso, MS, Yockteng, R (2023). Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity. *Front. Plant Sci.* 13, 1007104. doi: 10.3389/fpls.2022.1007104

Borowski, JM, Galli, V, da Silva Messias, R, Perin, EC, Buss, JH, dos Anjos e Silva, SD, et al. (2014). Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. *Planta* 239, 1187–1200. doi: 10.1007/s00425-014-2041-2

Bu, C, Zhang, Q, Zeng, J, Cao, X, Hao, Z, Qiao, D, et al. (2020). Identification of a novel anthocyanin synthesis pathway in the fungus *Aspergillus sydowii* H-1. *BMC Genomics* 21, 29. doi: 10.1186/s12864-019-6442-2

Cao, S, Sawettalake, N, Li, P, Fan, S, Shen, L (2024). DNA methylation variations underlie lettuce domestication and divergence. *Genome Biol.* 25, 158. doi: 10.1186/s13059-024-03310-x

Carravedo, M, Mallor, C, Garcés-Claver, A (2011). Descriptiva de las variedades de origen aragonés seleccionadas, in *Evaluación morfológica y molecular de variedades autóctonas aragonesas de lechuga (*Lactuca sativa *L.) y especies silvestres emparentadas (*Lactuca *spp.) conservadas en el banco de germoplasma de Especies Hortícolas de Zaragoza*, (Zaragoza: Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA) - Gobierno de Aragón), 101–170.

Casals Missio, J, Rivera, A, Figàs, MR, Casanova, C, Camí, B, Soler, S, et al. (2018). A Comparison of Landraces vs. Modern Varieties of Lettuce in Organic Farming During the Winter in the Mediterranean Area: An Approach Considering the Viewpoints of Breeders, Consumers, and Farmers. *Front. Plant Sci.* 9, 1491. doi: 10.3389/fpls.2018.01491

Castellarin, SD, Pfeiffer, A, Sivilotti, P, Degan, M, Peterlunger, E, Di Gaspero, G (2007). Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant, Cell Environ.* 30, 1381–1399. doi: 10.1111/j.1365-3040.2007.01716.x

Chalker-Scott, L (1999). Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochem. Photobiol.* 70, 1–9. doi: 10.1111/j.1751-1097.1999.tb01944.x

Chaves-Silva, S, dos Santos, AL, Chalfun-Júnior, A, Zhao, J, Peres, LEP, Benedito, VA (2018). Understanding the genetic regulation of anthocyanin biosynthesis in plants – Tools for breeding purple varieties of fruits and vegetables. *Phytochemistry* 153, 11–27. doi: 10.1016/j.phytochem.2018.05.013

Chen, W, Miao, Y, Ayyaz, A, Hannan, F, Huang, Q, Ulhassan, Z, et al. (2022). Purple stem *Brassica napus* exhibits higher photosynthetic efficiency, antioxidant potential and anthocyanin biosynthesis related genes expression against drought stress. *Front. Plant Sci.* 13, 936696. doi: 10.3389/fpls.2022.936696

Chen, Y, Li, T, Yang, Q, Zhang, Y, Zou, J, Bian, Z, et al. (2019). UVA Radiation is Beneficial for Yield and Quality of Indoor Cultivated Lettuce. *Front. Plant Sci.* 10, 1563. doi: 10.3389/fpls.2019.01563

Collard, BC, Mackill, DJ (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170

Damerum, A, Chapman, MA, Taylor, G (2020). Innovative breeding technologies in lettuce for improved post-harvest quality. *Postharvest Biol. Technol.* 168, 111266. doi: 10.1016/j.postharvbio.2020.111266

Damerum, A, Selmes, SL, Biggi, GF, Clarkson, GJJ, Rothwell, SD, Truco, MJ, et al. (2015). Elucidating the genetic basis of antioxidant status in lettuce (*Lactuca sativa*). *Hortic. Res.* 2, 15055. doi: 10.1038/hortres.2015.55

De Vries, IM (1997). Origin and domestication of *Lactuca sativa* L. *Genet. Resour. Crop Evol.* 44, 165–174. doi: 10.1023/A:1008611200727

Di Vita, G, Zanchini, R, Spina, D, Vastola, A, D'Amico, M, Caracciolo, F (2024). Simply red? The effects of distinct colours and sustainable production methods on the consumers' preferences for healthier sweet peppers. *Heliyon* 10, e28661. doi:

10.1016/j.heliyon.2024.e28661

Dyląg, A, Smoleń, S, Wisła-Świder, A, Kowalska, I, Sularz, O, Krzemińska, J, et al. (2023). Evaluation of the chemical composition and nutritional value of lettuce (*Lactuca sativa* L.) biofortified in hydroponics with iodine in the form of iodoquinolines. *Front. Plant Sci.* 14, 1288773. doi: 10.3389/fpls.2023.1288773

Everaert, C, Luypaert, M, Maag, JLV, Cheng, QX, DInger, ME, Hellemans, J, et al. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.* 7, 1559. doi: 10.1038/s41598-017-01617-3

FAO (2023). *The Impact of Disasters on Agriculture and Food Security 2023 - Avoiding and reducing losses through investment in resilience*. Rome, Italy. doi: 10.4060/cc7900en

FAO, IFAD, UNICEF, WFP, WHO (2023). *The State of Food Security and Nutrition in the World 2023. Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum*. Rome, Italy: FAO. doi: 10.4060/cc3017en

FAOSTAT (2022). Statistics of the Food and Agriculture Organization of the United Nations. Available at: https://www.fao.org/faostat/es/#data/QCL (Accessed May 5, 2024).

Farooq, M, Wahid, A, Kobayashi, N, Fujita, D, Basra, SMA (2009). Plant Drought Stress: Effects, Mechanisms and Management, in *Sustainable Agriculture*, eds. E. Lichtfouse, M. Navarrete, P. Debaeke, S. Véronique, and C. Alberola (Dordrecht: Springer), 153–188. doi: 10.1007/978-90-481-2666-8_12

Fath, MK, Naderi, M, Hamzavi, H, Ganji, M, Shabani, S, Ghahroodi, FN, et al. (2022). Molecular mechanisms and therapeutic effects of different vitamins and minerals in COVID-19 patients. *J. Trace Elem. Med. Biol.* 73, 127044. doi: 10.1016/j.jtemb.2022.127044

Funk, VA, Susanna, A, Stuessy, TF, Robinson, H (2009). Classification of Compositae, in *Systematics, Evolution, and Biogeography of Compositae*, eds. V. A. Funk, A. Susanna, T. Stuessy, and R. J. Bayer (Viena, Austria: IAPT), 171–189.

García-Macías, P, Ordidge, M, Vysini, E, Waroonphan, S, Battey, NH, Gordon, MH, et al. (2007). Changes in the Flavonoid and Phenolic Acid Contents and Antioxidant Activity of Red Leaf Lettuce (Lollo Rosso) Due to Cultivation under Plastic Films Varying in Ultraviolet Transparency. *J. Agric. Food Chem.* 55, 10168–10172. doi: 10.1021/jf071570m

Genesys (2024). Gateway to genetic resources. Available at: https://www.genesys-pgr.org/a/overview (Accessed May 14, 2024).

Guo, X, Liu, RH, Fu, X, Sun, X, Tang, K (2013). Over-expression of l-galactono-γ-

lactone dehydrogenase increases vitamin C, total phenolics and antioxidant activity in lettuce through bio-fortification. *Plant Cell. Tissue Organ Cult.* 114, 225–236. doi: 10.1007/s11240-013-0318-y

Guo, Z, Li, B, Du, J, Shen, F, Zhao, Y, Deng, Y, et al. (2023). LettuceGDB: The community database for lettuce genetics and omics. *Plant Commun.* 4, 100425. doi: 10.1016/j.xplc.2022.100425

Hamazaki, K, Kajiya-Kanegae, H, Yamasaki, M, Ebana, K, Yabe, S, Nakagawa, H, et al. (2020). Choosing the optimal population for a genome-wide association study: A simulation of whole-genome sequences from rice. *Plant Genome* 13, e20005. doi: 10.1002/tpg2.20005

Hao, JH, Zhang, LL, Li, PP, Sun, YC, Li, JK, Qin, XX, et al. (2018). Quantitative proteomics analysis of lettuce (*Lactuca sativa* L.) reveals molecular basis-associated auxin and photosynthesis with bolting induced by high temperature. *Int. J. Mol. Sci.* 19, 2967. doi: 10.3390/ijms19102967

Hassan, MN, Mekkawy, SA, Mahdy, M, Salem, KFM, Tawfik, E (2021). Recent molecular and breeding strategies in lettuce (*Lactuca* spp.). *Genet. Resour. Crop Evol.* 68, 3055–3079. doi: 10.1007/s10722-021-01246-w

Hinojosa-Gómez, J, San Martín-Hernández, C, Heredia, JB, León-Félix, J, Osuna-Enciso, T, Muy-Rangel, MD (2020). Anthocyanin Induction by Drought Stress in the Calyx of Roselle Cultivars. *Molecules* 25, 1555. doi: 10.3390/molecules25071555

Huo, H, Henry, IM, Coppoolse, ER, Verhoef-Post, M, Schut, JW, de Rooij, H, et al. (2016). Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing. *Plant J.* 88, 345–360. doi: 10.1111/tpj.13267

IPCC (2023). Summary for Policymakers, in *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, eds. Core Writing Team, H Lee, and J Romero. Geneva, Switzerland: IPCC. doi:10.59327/IPCC/AR6-9789291691647.001

Ishikawa, T, Dowdle, J, Smirnoff, N (2006). Progress in manipulating ascorbic acid biosynthesis and accumulation in plants. *Physiol. Plant.* 126, 343–355. doi: 10.1111/j.1399-3054.2006.00640.x

Jaffe, GM (1984). Vitamin C, in *Handbook of Vitamins*, ed. L. Machlin (New York: Marcel Dekker Inc), 199–244.

Jagger, IC, Whitaker, TW, Uselman, JJ, Owen, WM (1941). *The Imperial Strains of Lettuce*. U.S. Department of Agriculture.

Jansen, J, Verbakel, H, Peleman, J, van Hintum, TJL (2006). A note on the measurement

of genetic diversity within genebank accessions of lettuce (*Lactuca sativa* L.) using AFLP markers. *Theor. Appl. Genet.* 112, 554–561. doi: 10.1007/s00122-005-0162-5

Ju, Y, Yang, B, He, S, Tu, T, Min, Z, Fang, Y, et al. (2019). Anthocyanin accumulation and biosynthesis are modulated by regulated deficit irrigation in Cabernet Sauvignon (*Vitis vinifera* L.) grapes and wines. *Plant Physiol. Biochem.* 135, 469–479. doi: 10.1016/j.plaphy.2018.11.013

Kim, BK, Park, SY, Jeon, BY, Min, BW (2004). Metabolic Engineering Increased Vitamin C Levels in Lettuce by Overexpression of a L-Gulono-g-lactone Oxidase. *Hortic. Environ. Biotechnol.* 45, 16–20.

Kim, MJ, Moon, Y, Tou, JC, Mou, B, Waterland, NL (2016). Nutritional value, bioactive compounds and health benefits of lettuce (*Lactuca sativa* L.). *J. Food Compos. Anal.* 49, 19–34. doi: 10.1016/j.jfca.2016.03.004

Koyama, R, Itoh, H, Kimura, S, Morioka, A, Uno, Y (2012). Augmentation of Antioxidant Constituents by Drought Stress to Roots in Leafy Vegetables. *Horttechnology* 22, 121–125. doi: 10.21273/HORTTECH.22.1.121

Koyama, R, Yoshimoto, A, Ishibashi, M, Itoh, H, Uno, Y (2021). Enzymatic Activities and Gene Transcript Levels Associated with the Augmentation of Antioxidant Constituents during Drought Stress in Lettuce. *Horticulturae* 7, 444. doi: 10.3390/horticulturae7110444

Kwon, S-J, Truco, MJ, Hu, J (2012). LSGermOPA, a custom OPA of 384 EST-derived SNPs for high-throughput lettuce (*Lactuca sativa* L.) germplasm fingerprinting. *Mol. Breed.* 29, 887–901. doi: 10.1007/s11032-011-9623-5

Kwon, S, Simko, I, Hellier, B, Mou, B, Hu, J (2013). Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines. *Crop J.* 1, 25–33. doi: 10.1016/j.cj.2013.07.014

Landi, M, Tattini, M, Gould, KS (2015). Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* 119, 4–17. doi: 10.1016/j.envexpbot.2015.05.012

Landry, BS, Kesseli, R V, Farrara, B, Michelmore, RW (1987). A Genetic Map of Lettuce (*Lactuca sativa* L.) With Restriction Fragment Length Polymorphism, Isozyme, Disease Resistance and Morphological Markers. *Genetics* 116, 331–337. doi: 10.1093/genetics/116.2.331

Lebeda, A, Doležalová, I, Astley, D (2004). Representation of wild *Lactuca* spp. (Asteraceae, Lactuceae) in world genebank collections. *Genet. Resour. Crop Evol.* 51, 167–174. doi: 10.1023/B:GRES.0000020860.66075.f7

Lebeda, A, Křístková, E, Kitner, M, Mieslerová, B, Jemelková, M, Pink, DAC (2014).

Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. *Eur. J. Plant Pathol.* 138, 597–640. doi: 10.1007/s10658-013-0254-z

Lee, SK, Kader, AA (2000). Preharvest and postharvest factors influencing vitamin C content of horticultural crops. *Postharvest Biol. Technol.* 20, 207–220. doi: 10.1016/S0925-5214(00)00133-2

Li, Y, Schellhorn, HE (2007). New Developments and Novel Therapeutic Perspectives for Vitamin C. *J. Nutr.* 137, 2171–2184. doi: 10.1093/jn/137.10.2171

Lind, J (1753). *A Treatise of the Scurvy, in Three Parts. Containing an Inquiry into the Nature, Causes, and Cure, of that Disease*., eds. Sands, Murray, and Cochran. Edinburgh. doi: 10.1017/CBO9781107256644

Lindqvist, K (1960). On the origin of cultivated lettuce. *Hereditas* 46, 319–350. doi: 10.1111/j.1601-5223.1960.tb03091.x

Llorach, R, Martínez-Sánchez, A, Tomás-Barberán, FA, Gil, MI, Ferreres, F (2008). Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chem.* 108, 1028–1038. doi: 10.1016/j.foodchem.2007.11.032

Mallor, C, Bertolín, JR, Paracuellos, P, Juan, T (2023). Nutraceutical Potential of Leafy Vegetables Landraces at Microgreen, Baby, and Adult Stages of Development. *Foods* 12, 3173. doi: 10.3390/foods12173173

Mammana, I (2014). Concentration of market power in the EU seed market.

MAPA (2019). Informe del consumo de alimentación en España 2019. https://www.mapa.gob.es/en/alimentacion/temas/consumo-tendencias/informe2019_v2_tcm38-540250.pdf (Accessed May 5, 2024).

MAPA (2022). Informe del consumo de alimentación en España 2022. https://www.mapa.gob.es/va/alimentacion/temas/consumo-tendencias/informe-consumo-2022-baja-res_tcm39-655390.pdf (Accessed May 5, 2024).

Martínez-Ispizua, E, Calatayud, Á, Marsal, JI, Cannata, C, Basile, F, Abdelkhalik, A, et al. (2022). The Nutritional Quality Potential of Microgreens, Baby Leaves, and Adult Lettuce: An Underexploited Nutraceutical Source. *Foods* 11, 423. doi: 10.3390/foods11030423

Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2022). Validación de nuevos genes de referencia para estudios de expresión diferencial de genes involucrados en la síntesis de antocianinas en lechuga y especies silvestres relacionadas, in *Acta de Horticultura 90 (X Congreso Nacional de Mejora Genética de Plantas).* Eds. RA Malvar, P Fiz Rocha, (Pontevedra, Spain, Sociedad Española de Ciencias Hortícolas), 242-245.

Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2023). Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives. *Int. J. Mol. Sci.* 24, 3052. doi: 10.3390/ijms24033052

Medina-Lozano, I, Bertolín, JR, Díaz, A (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* 359, 129864. doi: 10.1016/j.foodchem.2021.129864

Medina-Lozano, I, Bertolín, JR, Díaz, A (2024a). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15, 3389. doi: 10.3389/fpls.2024.1369658

Medina-Lozano, I, Bertolín, JR, Plieske, J, Ganal, M, Gnad, H, Díaz, A (2024b). Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays. *Plant Genome* 17, e20518. doi: 10.1002/tpg2.20518

Medina-Lozano, I, Bertolín, JR, Zufiaurre, R, Díaz, A (2020). Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.). *J. Vis. Exp.* 160, e61440. doi: 10.3791/61440

Medina-Lozano, I, Díaz, A (2021). Nutritional Value and Phytochemical Content of Crop Landraces and Traditional Varieties, in *Landraces – Traditional Variety and Natural Breed*, ed. A. Elkelish (IntechOpen), 95–116. doi: 10.5772/intechopen.95514

Medina-Lozano, I, Díaz, A (2022a). Applications of Genomic Tools in Plant Breeding: Crop Biofortification. *Int. J. Mol. Sci.* 23, 3086. doi: 10.3390/ijms23063086

Medina-Lozano, I, Díaz, A (2022b). Evolución de la mejora genética en lechuga. *Agricultura* 1059, 86–89.

Medina-Lozano, I, Grimplet, J, Díaz, A (2025). Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress. *Front. Plant Sci.* 15, 1494339. doi: 10.3389/fpls.2024.1494339

Michelmore, R, Marsh, E, Seely, S, Landry, B (1987). Transformation of lettuce (*Lactuca sativa*) mediated by *Agrobacterium tumefaciens*. *Plant Cell Rep.* 6, 439–442. doi: 10.1007/BF00272777

Migliore, G, Galatia, A, Romeo, P, Crescimanno, M, Schifani, G (2015). Quality attributes of cactus pear fruit and their role in consumer choice: the case of Italian consumers. *Br. Food J.* 117, 1637–1651. doi: 10.1108/eb011695

Mikel, MA (2007). Genealogy of Contemporary North American Lettuce. *HortScience* 42, 489–493. doi: 10.21273/hortsci.42.3.489

Moreno-Escamilla, JO, Jimeńez-Hernández, FE, Alvarez-Parrilla, E, De La Rosa, LA,

Martínez-Ruiz, NDR, González-Fernández, R, et al. (2020). Effect of Elicitation on Polyphenol and Carotenoid Metabolism in Butterhead Lettuce (*Lactuca sativa* var. capitata). *ACS Omega* 5, 11535–11546. doi: 10.1021/acsomega.0c00680

Moreno-Vázquez, S, Ochoa, OE, Faber, N, Chao, S, Jacobs, JME, Maisonneuve, B, et al. (2003). SNP-based codominant markers for a recessive gene conferring resistance to corky root rot (*Rhizomonas suberifaciens*) in lettuce (*Lactuca sativa*). *Genome* 46, 1059–1069. doi: 10.1139/g03-073

Mou, B (2005). Genetic Variation of Beta-Carotene and Lutein Contents in Lettuce. *J. Am. Soc. Hortic. Sci.* 130, 870–876. doi: 10.21273/jashs.130.6.870

Mou, B (2008). Lettuce, in *Vegetables I. Asteraceae, brassicaceae, chenopodiaceae, and cucurbitaceae*, eds. J. Prohens and F. Nuez (New York: Springer Science), 75–116. doi: 10.1007/978-0-387-30443-4_3

Niyazova, NN, Huseynova, IM (2024). The Antioxidant Defense System of Tomato (*Solanum lycopersicum* L.) Varieties under Drought Stress and upon Post-Drought Rewatering. *Biochemistry* 89, 1146–1157. doi: 10.1134/S0006297924060130

OECD (2018). *Concentration in Seed Markets: Potential Effects and Policy Responses*. Paris: OECDPublishing. doi: 10.1787/9789264308367-en

Qaderi, MM, Martel, AB, Strugnell, CA (2023). Environmental Factors Regulate Plant Secondary Metabolites. *Plants* 12, 447. doi: 10.3390/plants12030447

Reyes-Chin-Wo, S, Wang, Z, Yang, X, Kozik, A, Arikit, S, Song, C, et al. (2017). Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8, 14953. doi: 10.1038/ncomms14953

Ruggieri, V, Sacco, A, Calafiore, R, Frusciante, L, Barone, A (2015). Dissecting a QTL into Candidate Genes Highlighted the Key Role of Pectinesterases in Regulating the Ascorbic Acid Content in Tomato Fruit. *Plant Genome* 8, 1–10. doi: 10.3835/plantgenome2014.08.0038

Rugienius, R, Bendokas, V, Siksnianas, T, Stanys, V, Sasnauskas, A, Kazanaviciute, V (2021). Characteristics of *Fragaria vesca* Yield Parameters and Anthocyanin Accumulation under Water Deficit Stress. *Plants* 10, 557. doi: 10.3390/plants10030557

Santos, FT, Goufo, P, Santos, C, Botelho, D, Fonseca, J, Queirós, A, et al. (2016). Comparison of five agro-industrial waste-based composts as growing media for lettuce: Effect on yield, phenolic compounds and Vitamin C. *Food Chem.* 209, 293–301. doi: 10.1016/j.foodchem.2016.04.087

Sarker, U, Oba, S (2018). Drought Stress Effects on Growth, ROS Markers, Compatible Solutes, Phenolics, Flavonoids, and Antioxidant Activity in *Amaranthus tricolor*.

Appl. Biochem. Biotechnol. 186, 999–1016. doi: 10.1007/s12010-018-2784-5

Sauvage, C, Segura, V, Bauchet, G, Stevens, R, Do, PT, Nikoloski, Z, et al. (2014). Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant Physiol.* 165, 1120–1132. doi: 10.1104/pp.114.241521

Seminario, A, Song, L, Zulet, A, Nguyen, HT, González, EM, Larrainzar, E (2017). Drought Stress Causes a Reduction in the Biosynthesis of Ascorbic Acid in Soybean Plants. *Front. Plant Sci.* 8, 1042. doi: 10.3389/fpls.2017.01042

Simko, I (2013). Marker-Assisted Selection for Disease Resistance in Lettuce, in *Translational Genomics for Crop Breeding, Volume I: Biotic Stress*, eds. R.K. Varshney and R. Tuberosa (John Wiley & Sons, Inc), 267–289. doi: 10.1002/9781118728475.ch14

Simko, I, Jia, M, Venkatesh, J, Kang, BC, Weng, Y, Barcaccia, G, et al. (2021). Genomics and Marker-Assisted Improvement of Vegetable Crops. *CRC. Crit. Rev. Plant Sci.* 40, 303–365. doi: 10.1080/07352689.2021.1941605

Simko, I, Pechenick, DA, McHale, LK, Truco, MJ, Ochoa, OE, Michelmore, RW, et al. (2009). Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. *BMC Plant Biol.* 9, 135. doi: 10.1186/1471-2229-9-135

Sng, BJR, Mun, B, Mohanty, B, Kim, M, Phua, ZW, Yang, H, et al. (2021). Combination of red and blue light induces anthocyanin and other secondary metabolite biosynthesis pathways in an age-dependent manner in Batavia lettuce. *Plant Sci.* 310, 110977. doi: 10.1016/j.plantsci.2021.110977

Su, W, Tao, R, Liu, W, Yu, C, Yue, Z, He, S, et al. (2020). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* 18, 479–490. doi: 10.1111/pbi.13213

Svirbely, JL, Szent-Györgyi, A (1932). Hexuronic Acid as the Antiscorbutic Factor. *Nature* 129, 576. doi: 10.1038/129576a0

Tanaka, Y, Sasaki, N, Ohmiya, A (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x

Tardin, FFD, Teixeira, A, Gonzaga, M, Celeste, M, Vidigal, G, Daher, RF, et al. (2003). Genetic diversity and determination of the optimum number of RAPD markers in lettuce (*Lactuca sativa* L.). *Acta Sci. Agron.* 25, 1–5.

Tripodi, P, Beretta, M, Peltier, D, Kalfas, I, Vasilikiotis, C, Laidet, A, et al. (2023). Development and application of Single Primer Enrichment Technology (SPET) SNP assay for population genomics analysis and candidate gene discovery in lettuce.

Front. Plant Sci. 14, 1252777. doi: 10.3389/fpls.2023.1252777

Truco, MJ, Ashrafi, H, Kozik, A, Van Leeuwen, H, Bowers, J, Wo, SRC, et al. (2013). An Ultra-High-Density, Transcript-Based, Genetic Map of Lettuce. *G3 Genes, Genomes, Genet.* 3, 617–631. doi: 10.1534/g3.112.004929

Tsormpatsidis, E, Henbest, RGC, Davis, FJ, Battey, NH, Hadley, P, Wagstaffe, A (2008). UV irradiance as a major influence on growth, development and secondary products of commercial importance in Lollo Rosso lettuce "Revolution" grown under polyethylene films. *Environ. Exp. Bot.* 63, 232–239. doi: 10.1016/j.envexpbot.2007.12.002

UPOV (2021). *Guidelines for the Conduct of Tests for Distinctness, Homogeneity, and Stability*. Geneva, Switzerland.

USDA (2022). FoodData Central. Available at: https://fdc.nal.usda.gov/

van Treuren, R, van Eekelen, HDLM, Wehrens, R, de Vos, RCH (2018). Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. *Metabolomics* 14, 146. doi: 10.1007/s11306-018-1443-8

Venkatesh, J, Park, SW (2014). Role of L-ascorbate in alleviating abiotic stresses in crop plants. *Bot. Stud.* 55, 38. doi: 10.1186/1999-3110-55-38

Viacava, GE, Roura, SI, Berrueta, LA, Iriondo, C, Gallo, B, Alonso-Salces, RM (2017). Characterization of phenolic compounds in green and red oak-leaf lettuce cultivars by UHPLC-DAD-ESI-QToF/MS using MS^E scan mode. *J. Mass Spectrom.* 52, 873–902. doi: 10.1002/jms.4021

Wada, KC, Inagaki, N, Sakai, H, Yamashita, H, Nakai, Y, Fujimoto, Z, et al. (2022). Genetic effects of *Red Lettuce Leaf* genes on red coloration in leaf lettuce under artificial lighting conditions. *Plant-Environment Interact.* 3, 179–192. doi: 10.1002/pei3.10089

Wei, T, van Treuren, R, Liu, X, Zhang, Z, Chen, J, Liu, Y, et al. (2021). Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* 53, 752–760. doi: 10.1038/s41588-021-00831-0

Wheeler, GL, Jones, MA, Smirnoff, N (1998). The biosynthetic pathway of vitamin C in higher plants. *Nature* 393, 365–369. doi: 10.1038/30728

Woo, JW, Kim, J, Kwon, SI, Corvalán, C, Cho, SW, Kim, H, et al. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* 33, 1162–1164. doi: 10.1038/nbt.3389

Wu, X, Prior, RL (2005). Identification and characterization of anthocyanins by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry in common foods in the United States: vegetables, nuts, and grains. *J. Agric. Food Chem.* 53, 3101–3113. doi: 10.1021/jf0478861

Yim, AKY, Wong, JWH, Ku, YS, Qin, H, Chan, TF, Lam, HM (2015). Using RNA-seq Data to Evaluate Reference Genes Suitable for Gene Expression Studies in Soybean. *PLoS One* 10, e0136343. doi: 10.1371/journal.pone.0136343

Yousuf, B, Gul, K, Wani, AA, Singh, P (2016). Health Benefits of Anthocyanins and Their Encapsulation for Potential Use in Food Systems: a Review. *Crit. Rev. Food Sci. Nutr.* 56, 2223–2230. doi: 10.1080/10408398.2013.805316

Zeljković, SC, Štefelová, N, Hron, K, Doležalová, I, Tarkowski, P (2023). Preharvest Abiotic Stress Affects the Nutritional Value of Lettuce. *Agronomy* 13, 398. doi: 10.3390/ agronomy13020398

Zhang, H, Si, X, Ji, X, Fan, R, Liu, J, Chen, K, et al. (2018a). Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.* 36, 894–900. doi: 10.1038/nbt.4202

Zhang, L, Su, W, Tao, R, Zhang, W, Chen, J, Wu, P, et al. (2017). RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat. Commun.* 8, 2264. doi: 10.1038/s41467-017-02445-9

Zhang, W, Alseekh, S, Zhu, X, Zhang, Q, Fernie, AR, Kuang, H, et al. (2020). Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *Plant J.* 104, 613–630. doi: 10.1111/tpj.14950

Zhang, Y, Xu, S, Cheng, Y, Peng, Z, Han, J (2018b). Transcriptome profiling of anthocyaninrelated genes reveals effects of light intensity on anthocyanin biosynthesis in red leaf lettuce. *PeerJ* 13, e4607. doi: 10.7717/peerj.4607

Zhang, YZ, Xu, SZ, Cheng, YW, Ya, HY, Han, JM (2016). Transcriptome analysis and anthocyanin-related genes in red leaf lettuce. *Genet. Mol. Res.* 15, gmr.15017023. doi: 10.4238/gmr.15017023

Zhou, H, Yu, L, Liu, S, Zhu, A, Yang, Y, Chen, C, et al. (2023). Transcriptome comparison analyses in UV-B induced AsA accumulation of *Lactuca sativa* L. *BMC Genomics* 24, 61. doi: 10.1186/s12864-023-09133-7

Zhou, SM, Kong, XZ, Kang, HH, Sun, XD, Wang, W (2015). The Involvement of Wheat F-box Protein Gene *TaFBA1* in the Oxidative Stress Tolerance of Plants. *PLoS One* 10, e0122117. doi: 10.1371/journal.pone.0122117

Zhou, WL, Liu, WK, Yang, QC (2007). Quality changes in hydroponic lettuce grown under pre-harvest short-duration continuous light of different intensities. *J. Hortic. Sci. Biotechnol.* 87, 429–434. doi: 10.1080/14620316.2012.11512890

Zhou, Z, Cong, P, Tian, Y, Zhu, Y (2017). Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. *PLoS One* 12, e0185288. doi: 10.1371/journal.pone.0185288

ANNEXES

ANNEX 1

Medina-Lozano, I, Díaz, A (2022b). Evolución de la mejora genética en lechuga. *Agricultura* 1059, 86–89.

Evolución de la mejora genética en lechuga

La mejora de la lechuga comenzó hace más de 10.000 años con el nacimiento de la agricultura. La domesticación de los ancestros silvestres generó distintas variedades de lechuga cultivada que se han diversificado gracias a los programas de mejora de épocas más recientes. Estos han empleado herramientas que van desde la realización de cruzamientos dirigidos hasta la edición genómica.

Inés Medina-Lozano[1,2], Aurora Díaz[1,2]
[1] Departamento de Ciencia Vegetal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Zaragoza, España.
[2] Instituto Agroalimentario de Aragón – IA2 (CITA-Universidad de Zaragoza), Zaragoza, España

La lechuga (*Lactuca sativa* L.) es una de las hortalizas de hoja más importantes a nivel mundial. De hecho, su producción total junto con la de endibia superó los 27 millones de toneladas en el año 2020 (FAOSTAT, 2020).

Según la UPOV (Unión para la Protección de Nuevas Variedades de Plantas), la lechuga se puede clasificar en función de su morfología en 12 tipos distintos: Batavia, Frisée d'Amérique, Gem, Hoja de Roble, Iceberg, Lollo, Mantecosa, Multi-dividida, Novita, Rizada, Romana y de Tallo. En función del color de la hoja, podemos encontrar tres tipos: verde, roja y semi-roja. Dentro del género *Lactuca* también podemos encontrar una gran variabilidad, pues cuenta con más de 100 especies distintas, habiendo hasta 20 en el caso del acervo genético primario (más estrechamente relacionado) donde se engloba la especie cultivada.

Los bancos de germoplasma juegan un papel fundamental en la conservación y la caracterización de todo este material vegetal. En el caso de la lechuga, el Centro de Recursos Genéticos de los Países Bajos (CGN, del inglés Centre for Genetic Resources (Wageningen, Netherlands)) cuenta con la colección más completa, con más de 2.500 accesiones, pero existen muchos más bancos de germoplasma en el mundo, donde no solo se conservan variedades modernas y especies silvestres, sino también variedades tradicionales del cultivo **(Figura 1)**.

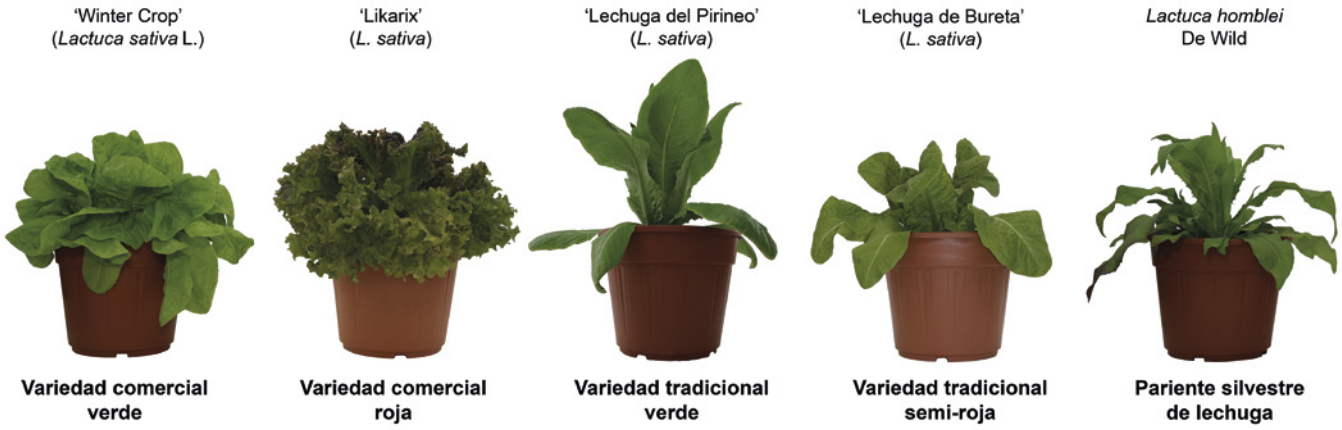
Inicios de la mejora en lechuga

La mejora de plantas ha existido desde que comenzó su domesticación, es decir, con el origen de la agricultura, unos 10.000 años atrás. En el caso de la lechuga, se cree que el abanico de variedades cultivadas existentes hoy día surgió a partir de la domesticación de su pariente silvestre *Lactuca serriola* L., aunque no se descartan otras hipótesis y es bastante probable que más especies silvestres hayan también contribuido en mayor o menor medida. Además, un estudio reciente en el que se han secuenciado 445 accesiones de *Lactuca* ha planteado que esta domesticación pudo haber comenzado hace unos 6.000 años alrededor de la región del Cáucaso (Wei *et al.*, 2021). Hacia el año 2.500 a.C. se iniciaría el cultivo de una lechuga primitiva para la extracción del aceite de sus semillas en Egipto. Posteriormente, comenzaría la transición del cultivo de variedades de semilla oleaginosa hacia variedades de hoja comestible, que se cree surgieron en Grecia y Roma. Ya en épocas más recientes, se fueron obteniendo variedades más modernas que aparecieron primero en el resto del continente europeo y más adelante en Estados Unidos. Se sabe que en estas últimas hubo introgresión de genes de otra especie silvestre emparentada con la cultivada, *Lactuca virosa* L. Desde el Cáucaso también podría haberse producido la expansión del cultivo hacia China, donde predominan las variedades de tallo **(Figura 2)**.

La domesticación de las especies silvestres de *Lactuca* conllevó importantes cambios en distintos caracteres y dio lugar a la lechuga cultivada tal y como la conocemos hoy en día. Por ejemplo, se produjo la pérdida de espinas en hojas y tallos, una reducción del sabor amargo y de la cantidad de látex (estos dos últimos estrechamente relacionados) y el retraso de la floración. También provocó un aumento en el tamaño de las semillas y cambios en la forma de las hojas y se indujo la formación de cogollos en algunas variedades.

Es importante destacar que la lechuga cultivada se reproduce normalmente mediante autofecundación. Es decir, de forma natural la especie constituye líneas puras, caracterizadas por una alta uniformidad fenotípica y genotípica, debido a un alto grado de homocigosidad **(Figura 3)**. Sin embargo, se cree que los ancestros de los que deriva son

Figura 1
Variedades comerciales de lechuga de hoja verde y roja; variedades tradicionales de hoja verde y semi-roja; especie silvestre emparentada con la lechuga



alógamos, sobre todo los más cercanos al centro de origen, lo que facilitaría los intercambios genéticos (introgresión de genes) entre las distintas especies y, por tanto, explicaría la diversidad de tipos de lechuga cultivada existente.

Evolución de la mejora

- Mejora convencional

Con la puesta en práctica de la mejora convencional, la obtención de nuevas variedades se basó en el cruzamiento dirigido entre plantas. Para ello, es común llevar a cabo retrocruzamientos sucesivos empleando plantas donantes con los caracteres de interés a introducir y plantas receptoras (parental recurrente) con características agronómicas apropiadas para su cultivo en los sistemas agrícolas de la época. Hasta el momento, los programas de mejora de lechuga se han centrado fundamentalmente en la incorporación de resistencias a enfermedades, aunque también en la adquisición de tolerancia a estreses abióticos y el retraso en la producción de flores, entre otros aspectos.

El primer programa de mejora de lechuga que se conoce lo inició el Departamento de Agricultura de Estados Unidos en el año 1922 y tenía como objetivo el control de un tizón de origen fúngico. Con este programa se logró desarrollar una variedad Iceberg resistente ('Imperial') mediante cruces entre variedades de tipo Iceberg, Romana y Mantecosa. Desde entonces, se han emprendido muchos

Figura 2
Mapa de la posible domesticación y dispersión de la lechuga cultivada.

otros programas de mejora en los que los cruzamientos entre distintas variedades cultivadas han posibilitado la obtención de muchas otras (Mikel, 2007).

Como se ha indicado anteriormente, la lechuga es una especie de naturaleza autógama y esto supone una limitación importante para los programas de mejora. La autogamia provoca una reducción de la diversidad genética presente en la especie, lo que supone un gran inconveniente para la mejora convencional de plantas. Además, la falta de diversidad es mayor en las variedades comerciales, especialmente en las de tipo Iceberg. Por lo tanto, este inconveniente se podría superar en cierta medida realizando cruces con variedades tradicionales o con los parientes silvestres de la lechuga, que han sido sometidos a menor o a ninguna presión de mejora, respectivamente, y exhiben por tanto una mayor variabilidad genética. Los parientes silvestres representan unos recursos genéticos muy valiosos y, en el caso de la mejora de la lechuga, los más explotados han sido *L. serriola*, *Lactuca saligna* L. y *L. virosa* (Lebeda et al., 2014). De nuevo, su uso principal ha ido dirigido a la incorporación de resistencias a enfermedades. Por ejemplo, *L. serriola* y *L. saligna* se han empleado en múltiples programas de mejora para transferir resistencia a mildiu, una de las enfermedades fúngicas más importantes de la lechuga a nivel mundial. En el caso de *L. virosa*, se necesitan cruces previos con *L. serriola* porque se trata de un pariente más lejano, pero también se ha empleado con éxito en la obtención

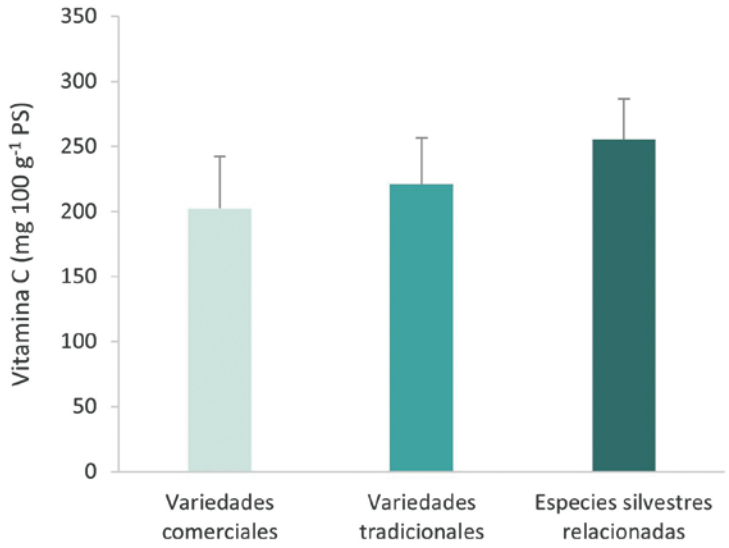
Figura 3
Plantas de la variedad tradicional 'Lechuga del Pirineo', muy uniforme tanto fenotípica como genéticamente.

de resistencia a mildiu. Además se han realizado cruzamientos con una lechuga silvestre egipcia que han permitido la incorporación de la resistencia al virus del mosaico de la lechuga en la variedad de tipo Iceberg 'Vanguardia'.
La calidad nutricional es otro aspecto a tener en cuenta en los programas de mejora. Se sabe que algunas variedades tradicionales y especies silvestres relacionadas son más ricas en algunos nutrientes y compuestos bioactivos (Medina-Lozano *et al.*, 2021). En la **Figura 4** se puede ver cómo las variedades tradicionales de lechuga y algunos de sus parientes silvestres evaluados en nuestro grupo tienen, de media, un mayor contenido en vitamina C. Sin embargo, en los programas de mejora convencional el valor nutricional normalmente se ha dejado de lado.
Por otro lado, el uso de las herramientas genómicas es una gran ayuda para los programas de mejora convencional. En la bibliografía se encuentra publicado un gran número de marcadores moleculares de lechuga que se han empleado con distintos objetivos, como la identificación de variedades, la construcción de mapas de ligamiento y la selección asistida por marcadores (MAS). Algunos ejemplos de estos marcadores son Polimorfismos en la Longitud de los Fragmentos de Restricción (RFLPs), ADN Polimórfico Amplificado al Azar (RAPDs), Polimorfismos en la Longitud de Fragmentos Amplificados (AFLPs), Repeticiones de Secuencia Simple (SSRs) y Polimorfismos de un Solo Nucleótido (SNPs), siendo estos últimos los más ampliamente utilizados hoy en día. De hecho, en la base de datos LettuceGBD (https://www.lettucegdb.com/) hay publicada una gran lista de SNPs en lechuga. Los mapas genéticos se han construido y actualizado en función de la disponibilidad de los marcadores moleculares. El primer mapa genético de lechuga constaba de 53 marcadores (Landry *et al.*, 1987), mientras que en el año 2013 se ha publicado un mapa de ultra alta densidad con 14.000 (Truco *et al.*, 2013). Más recientemente se ha secuenciado el genoma de *L. sativa* cv. Salinas (Reyes-Chin-Wo *et al.*, 2017). Toda esta información es fundamental en los programas de mejora para explotar de forma más eficiente la variabilidad genética disponible y para reducir de manera importante el tiempo y los costes. Asimismo, muchos estudios han encontrado marcadores asociados a genes o *loci* que controlan caracteres cuantitativos (QTL, del inglés Quantitative Trait Loci), por ejemplo, responsables de la resistencia a enfermedades fúngicas, como el mildiu y la marchitez causada por *Fusarium* o *Verticillium*, víricas, como las provocadas por el virus del mosaico del nabo, y bacterianas, como la mancha foliar bacteriana (Damerum *et al.*, 2020). También se han descrito QTL asociados a la tolerancia a estrés hídrico y salino, a caracteres agronómicos, como la vida útil, la longevidad de las semillas, la coloración de las hojas y el tiempo de floración, y al contenido en compuestos saludables, como antioxidantes (Damerum *et al.*, 2020). Los polimorfismos presentes en los genes subyacentes a estos QTL serían marcadores idóneos en estrategias de MAS.
Un ejemplo aplicado del uso de herramientas genómicas en lechuga es el empleo de marcadores moleculares en la obtención de nuevas variedades, como es el caso de una lechuga tipo Romana con una mayor vida útil y resistente a la enfermedad vírica conocida como muerte súbita (Simko *et al.*, 2010).

- Mejora genética moderna

En los últimos años la mejora genética moderna ha cobrado un enorme peso en los programas de mejora de cultivos. La transformación genética de plantas permite la transferencia de uno o más genes independientemente de su origen en términos de evolución y taxonomía, siendo posible incluso la transferencia entre reinos. De esta forma, se supera el inconveniente de la escasa variabilidad genética presente en las formas cultivadas, propio de la mejora convencional y, además, se consigue acortar en gran medida los tiempos de los programas de mejora.
El método más usado para la obtención de lechuga transgénica es la transferencia de genes mediada por *Agrobacterium*

Figura 4
Contenido medio de vitamina C (mg 100 g^{-1} peso seco (PS)) en variedades comerciales y tradicionales y en algunas especies silvestres relacionadas. Las barras representan la desviación estándar del total.



tumefaciens. En las últimas dos décadas se han obtenido lechugas transgénicas resistentes a hongos y virus (Lebeda *et al.*, 2014; Song *et al.*, 2014), pero también otros caracteres han sido objetivos importantes en este tipo de estrategias de mejora, como es el caso de la calidad nutricional. Por ejemplo, se han obtenido lechugas transgénicas con un mayor contenido en β-caroteno (vitamina A) y ácido fólico (vitamina B) (Fu *et al.*, 2012), ácido ascórbico (vitamina C) (Landi *et al.*, 2015) y tocoferoles (vitamina E) (Yabuta *et al.*, 2013). Por otro lado, también se ha empleado la transgénesis en lechuga para la producción de vacunas orales, para el aumento de la tolerancia al estrés por sequía y por frío y para el retraso de la floración (revisado en Lebeda *et al.*, 2014 y en Song *et al.*, 2014).

Uno de los últimos avances en la mejora genética de plantas es el empleo de la edición genómica. Con esta tecnología es posible conseguir una manipulación precisa de los genomas gracias al uso de nucleasas dirigidas. Actualmente, CRISPR-Cas9 es una de las técnicas de edición genómica más innovadoras y ya se ha aplicado en lechuga. Con el noqueo o silenciamiento mediante CRISPR-Cas9 de *LsNCED4*, gen que regula la termo-inhibición de la germinación de las semillas, se ha conseguido un gran aumento en la temperatura máxima de germinación en las variedades 'Salinas' y 'Cobham Green' (Bertier *et al.*, 2018). También se ha demostrado la eficacia de esta técnica en la obtención de plantas con hojas variegadas en la variedad 'Salinas' mediante el noqueo del gen *LsVAR2* (Nguyen et *al.*, 2021).

A pesar de sus ventajas, tanto los cultivos transgénicos como los obtenidos mediante edición genómica están sujetos a barreras legislativas muy prohibitivas, especialmente en Europa, donde solo está permitido el maíz Bt desarrollado por Monsanto, capaz de sintetizar la proteína Cry, producida naturalmente por *Bacillus thuringiensis* y que actúa como insecticida frente a las larvas de los insectos barrenadores del tallo. Sin embargo, la edición genómica tiene un futuro muy prometedor, ya que puede producir el noqueo de genes sin necesidad de introducir transgenes (ADN foráneo) en el producto final. De hecho, en algunos países, como Estados Unidos, Brasil y Canadá, los cultivos obtenidos mediante edición genómica ya se regulan de forma distinta a los transgénicos. Por tanto, pronto podrían aparecer en el mercado lechugas modificadas mediante ingeniería genética. Irónicamente, dichas lechugas podían ser comercializadas en países donde su cultivo no está permitido, como es el caso de España.

Conclusiones

Cabe destacar que la innovación en el campo de la mejora de la lechuga es constante. Por ejemplo, se han desarrollado recientemente variedades que permiten cubrir la producción durante todo el año y que a la vez están adaptadas a distintas zonas de cultivo. Otra tendencia actual en la mejora es la producción de lechugas baby, probablemente muy demandadas porque las familias son cada vez de menos miembros. Puede que en el futuro se aborden nuevos objetivos de mejora que hoy en día ni siquiera imaginamos. Su alcance será viable gracias a los avances tecnológicos y a una posible, e incluso necesaria, integración de las distintas estrategias de mejora convencionales y modernas.

Agradecimientos

Agradecemos al Banco de Germoplasma de Especies Hortícolas de Zaragoza (BGHZ-CITA, España) y al Centre for Genetic Resources (CGN , Wageningen, Países Bajos) por suministrar las semillas necesarias para este trabajo. Agradecemos a Juan Ramón Bertolín del "laboratorio de valoración nutritiva" del CITA por el apoyo técnico.
Este trabajo ha sido financiado por los proyectos RTA2017-00093-00-00 del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) y LMP164_18 del Gobierno de Aragón; y por el Programa Operativo FEDER Aragón 2014-2020 y el Fondo Social Europeo de la Unión Europea (Grupo Consolidado A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética")
I. M. L. ha contado con el apoyo de un contrato predoctoral para la formación de doctores del Ministerio de Ciencia, Innovación y Universidades de España (MCIU) y la Agencia Estatal de Investigación (AEI).

Bibliografía

Queda a disposición del lector interesado en el correo electrónico: redaccion@editorialagricola.com

ANNEX 2

Medina-Lozano, I., Díaz, A (2022a). Applications of Genomic Tools in Plant Breeding: Crop Biofortification. *Int. J. Mol. Sci.* 23, 3086. https://doi.org/10.3390/ijms23063086.

International Journal of
Molecular Sciences

MDPI

Review

Applications of Genomic Tools in Plant Breeding: Crop Biofortification

Inés Medina-Lozano [1,2] and Aurora Díaz [1,2,*]

[1] Departmento de Ciencia Vegetal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Universidad de Zaragoza, Avda. Montañana 930, 50059 Zaragoza, Spain; imedina@cita-aragon.es
[2] Instituto Agroalimentario de Aragón — IA2, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Universidad de Zaragoza, 50013 Zaragoza, Spain
* Correspondence: adiazb@cita-aragon.es; Tel.: +34-9-7671-6526 (ext. 816526)

Abstract: Crop breeding has mainly been focused on increasing productivity, either directly or by decreasing the losses caused by biotic and abiotic stresses (that is, incorporating resistance to diseases and enhancing tolerance to adverse conditions, respectively). Quite the opposite, little attention has been paid to improve the nutritional value of crops. It has not been until recently that crop biofortification has become an objective within breeding programs, through either conventional methods or genetic engineering. There are many steps along this long path, from the initial evaluation of germplasm for the content of nutrients and health-promoting compounds to the development of biofortified varieties, with the available and future genomic tools assisting scientists and breeders in reaching their objectives as well as speeding up the process. This review offers a compendium of the genomic technologies used to explore and create biodiversity, to associate the traits of interest to the genome, and to transfer the genomic regions responsible for the desirable characteristics into potential new varieties. Finally, a glimpse of future perspectives and challenges in this emerging area is offered by taking the present scenario and the slow progress of the regulatory framework as the starting point.

Keywords: biofortification; breeding; crop; cisgenesis; intragenesis; metabolic GWAS (mGWAS); single-nucleotide polymorphisms (SNPs); trangenesis

Citation: Medina-Lozano, I.; Díaz, A. Applications of Genomic Tools in Plant Breeding: Crop Biofortification. *Int. J. Mol. Sci.* **2022**, *23*, 3086. https://doi.org/10.3390/ijms23063086

Academic Editor: Frank M. You

Received: 31 January 2022
Accepted: 10 March 2022
Published: 13 March 2022

1. Introduction

Malnutrition is known to be a global public health problem and it has worsened with the COVID-19 pandemic. In 2020, about 768 million people in the world faced hunger, around 118 million more than in 2019 [1]. In addition, around 2.37 billion people (nearly one in three people in the world) suffered food insecurity (i.e., an inadequate access to safe, nutritious and sufficient food) in 2020, almost 320 million people more in just one year [1]. In fact, it is the first time that food insecurity has increased in North America and Europe since 2014 [1]. However, malnutrition is not only caused by the lack of food but also by a low dietary intake of essential nutrients (micronutrients included), known as hidden hunger [2]. This problem affects mainly developing countries in which the diet is usually based on more affordable major staple crops, characterized by a low micronutrient content. That being true, malnutrition is also present in developed countries, although in this case it is possibly due to unhealthy habits, such as extreme weight loss diets or substance abuse. It does not alleviate this situation given the fact that crop breeding has been mainly focused on increasing production, incorporating resistance to diseases, and enhancing tolerance to abiotic stresses, which has resulted in commercial varieties with low nutritional value [3].

Biofortification, i.e., the development of food crops with a high nutritional value per se through both conventional breeding and modern biotechnology techniques, could help in preventing hidden hunger. Micronutrients, minerals [4–12], vitamins [13–28],

or both [29], are the most common nutritional targets for biofortification strategies, though the improvement in fatty acid composition [30–33] and the increase in essential amino acids [34–37] and antioxidants [38–41] have also been recently included as aims of biofortification programs. This strategy carries multiple advantages. For example, it is a cost-effective approach, as shown by studies that report that for every dollar invested in the development of biofortified crops, as much as USD 17 of benefits may be obtained [42]. This is because, after a one-time investment to obtain the biofortified crops, they are able to synthesize larger amounts of the particular compounds without the need of adding any external micronutrients (fertilizers), which was the case in classical fortification. Therefore, as well as economic benefits, biofortification also brings environmental benefits. Moreover, it seems that breeding for a higher content in micronutrients does not entail a yield penalty [43,44]. This could be really helpful in developing countries, especially in areas with a limited access to marketed crops, as farmers could grow biofortified crops in the same way as conventional crops. Consequently, biofortification could be considered a sustainable and long-term solution to hidden hunger. In fact, the expected increase in population up to 9.7 billion by 2050 [45] makes it even more necessary.

Nevertheless, since the biofortification of a crop is tackled until the product is released to the market, a series of key steps have to be taken. The first would be to choose the species and the micronutrient to be enhanced. To maximize the positive impact on society, most consumed crops should be the target. This is what has been actually happening as, among the biofortified crops already developed, we can find staple crops, such as cereals (barley, maize, rice, and wheat) and beans, and some of the most consumed vegetables (tomato and potato) and fruits (apple and banana). One of the first steps consists of an evaluation of germplasm for their content in nutrients and health-promoting compounds; thus, outstanding alleles for those metabolic traits can be selected. Alternatively, the variability can be generated through induced mutagenesis (widely used in plant breeding since optimized during the second half of the 20th century) or by other more modern techniques of gene editing (i.e., clustered regularly interspaced short palindromic repeats (CRISPR)-associated system (CRISPR/Cas)). Secondly, genetic studies are usually conducted and molecular markers have to be developed to associate the trait of interest to the genomic regions. Finally, the allelic variants responsible for an increased content of the particular phytochemical have to be introduced to obtain the biofortified crop, either by conventional breeding or by modern biotechnology techniques. In this review, we will describe these steps in depth and, within the modern methods to introduce the allelic variants responsible for the increase in the specific compound, we will focus on transgenesis, cisgenesis, and intragenesis. Other simultaneous efforts will have to be made in order to ensure success both in the commercialization of the biofortified product and in the impact on consumers' health. For the first goal, studies of market potential and consumers' behavior and acceptability will have to be undertaken in advance, as was the case of selenium-biofortified apples [46] and iodine-biofortified fruits and vegetables [47], for example, both in Germany. This point is especially important in the case of controversial goods, such as transgenic biofortified food. That should be accompanied by promotion campaigns to make the product's beneficial properties public, as the one carried out with the orange-flesh sweet potato biofortified in pro-vitamin A in Ghana and Nigeria [48]. For the second objective, analyses of micronutrient bioavailability and their efficacy of conversion in the human body will have to be performed, as reported in intervention studies which supply vitamin A-biofortified maize to Zambian children with promising results [49].

Taking all the above into account, the present review aims, firstly, to summarize the genomic tools available to explore the variability through single-nucleotide polymorphism (SNP) genotyping, and the analytical methods to determine the phytochemical profile and/or content of plant food. Secondly, a compendium of the researches carried out on the genomic association of metabolic data in crops is also presented here. Thirdly, different methods used to transfer the genomic regions responsible for a raise in the compound

synthesis to the crops in order to create new biofortified varieties are shown, as well as some examples of their applications. These methods are either encompassed in conventional breeding strategies or modern biotechnology approaches, such as transgenesis, cisgenesis, and intragenesis. Finally, an overview of the current regulation and the future prospects of developing nutritionally enriched crops is also offered.

All the information needed to deal with the subjects mentioned above is obtained through searches in public databases and webpages, as described in Supplementary File S1.

2. Exploring Biodiversity: Searching for Outstanding Material

2.1. Genomic Diversity Enquired by SNP Genotyping

SNPs are not only the most frequent sequence variations among all practically genomes [50], but also the most amenable to automation. Even if a long list of molecular markers, and, more specifically, genetic markers, has been used in plant breeding since the 1980's (restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs), inter-simple sequence repeats (ISSRs), cleaved amplified polymorphic sequences (CAPS), etc.) [51], all of them have been unarguably ousted by SNPs. Their predominance is also a consequence of the development of next-generation sequencing (NGS), including second- and third-generation sequencing (SGS and TGS), mainly SGS, which evolved from the sequencing of short DNA fragments (first-generation sequencing, FGS) to high-throughput technologies (SGS) and, finally, single-molecule sequencing (TGS). This soon made necessary high-throughput SNP genotyping platforms that could produce a massive volume of data more cost-effectively in a short period of time. Among the wide variety of techniques developed to genotype SNPs and the different detection methods coupled to them, we will highlight those more commonly used nowadays with crops and those that process a medium (normally, in the laboratory) to high number of markers and samples (commercial platforms). All of them are based on hybridization, amplification, sequencing, or a combination of them, and they have been grouped according to the type of platform employed (Figure 1).

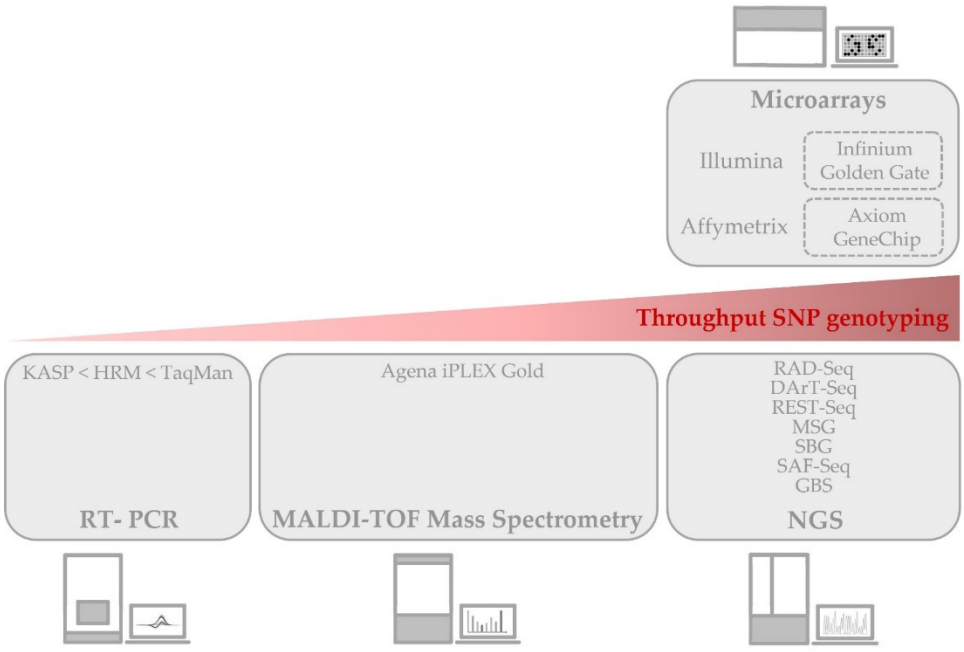


Figure 1. Comparison of the SNP genotyping techniques most commonly used in crops grouped by the platforms in the throughput level.

2.1.1. SNP Genotyping Microarrays

Among the assays available, the Affymetrix (Axiom) is a hybridization-based microarray that uses probes for both alleles. Independent of the allele at the particular locus, both probes hybridize with the DNA sample though the signal become dimmer in the case of a mismatch. So, the genotype of each SNP marker is called by the probes, showing the highest intensity in their signal. SNP Affymetrix arrays (either Axiom or GeneChip) have been used in a number of food crops, including cereals (maize [52,53], rice [54,55], rye [56], and wheat [57,58]), horticultural crops (chickpea [59], lettuce [60], potato [61], soybean [62], and strawberry [63]), and woody crops (apple tree [64] and peanut tree [65]), among others.

In the Illumina BeadArray (Infinium), the silica beads are coated with probes targeting a specific SNP locus. They bind the region just upstream the polymorphic site. Then, by single-base extension (SBE), a labelled nucleotide will be incorporated, emitting a different signal depending on the base. Illumina developed other BeadArray (GoldenGate) that uses fluorescent universal primers that hybridize to the allele-specific oligos. These technologies have been extensively used to discover and genotype SNPs in food crops, including cereals (barley [66], maize [67,68], oat [69], rice [70,71], and wheat [72]), oil crops (oilseed rape [73] and sunflower [74,75]), horticultural crops (cowpea [76], potato [77], tomato [78], and soybean [79]), and woody crops (apple tree [80,81], cherry tee [82], peach tree [83–85], pear tree [86], and vine [87,88]), among others.

The immobilization of samples, probes, ddNTP, etc. on chips (depending on the technique) is what makes interrogating hundreds of thousands or even millions of markers simultaneously feasible (Figure 1). In both cases, there are predesigned chips for some crops, which is the most affordable choice, but there is also the possibility of designing custom chips with the SNP markers of interest.

2.1.2. Real-Time PCR for SNP Genotyping

One of the commercially available assays within this category is the TaqMan SNP genotyping. This technology is also based in DNA hybridization and amplification, the signal is generated by fluorescence resonance energy transfer (FRET), and it is amenable to automation by real-time PCR though it does not reach the same high-throughput format than microarrays (Figure 1). Briefly, two allele-specific probes are designed for each SNP locus with two different fluorescent dyes attached to them. When the probe is free, the fluorescence is suppressed by quenching. Only when the probe perfectly hybridizes with the DNA fragment containing the SNP allele and is extended by PCR, the fluorophore is released by the exonuclease activity of the DNA polymerase and its signal is captured by the appropriate detector. These techniques have been mainly used in plants to diagnose pathogens and, to a smaller extent, to identify transgenes and detect food frauds, though there are also some cases where they are used to study the genetics behind some traits of interest in food crops, such as the presence of anthocyanins in potato skin [89].

As the previous one, the Kompetitive allele-specific PCR (KASP) is also a FRET method that makes use of hybridization and amplification though, unlike the TaqMan assay, the reagents for the allele-specific amplification, on the one hand, and the dye and quenchers, on the other, act in two phases. During a first round of PCR, the allele-specific and the common reverse primer amplifies the region by harboring the target SNP. After this, one of the fluor-labelled oligo that was quenched until now binds as a tail to the corresponding amplified allele, generating a fluorescent signal. KASP assays have been extensively used in different crops, mainly cereals, becoming very helpful for MAS in wheat [90–99], barley [100], rice [101–104], sorghum [105], pea [106], watermelon [107,108], faba bean [109], tomato [110,111], and *Brassica oleracea* (cabbage, broccoli, kohlrabi, and Chinese kale [112]).

Another methodology included here is the high-resolution melting (HRM) analysis. After the amplification by PCR of the region containing the SNP of interest in the presence of a dye that binds to double-stranded DNA, the products are melted into a single strand. This then causes the release of the dye and a decrease in its fluorescence. The real-time PCR is able to detect those changes and generate a melt curve that is different for each of

the genotypes at the SNP locus. Apart from cultivar identification, species authentication and pathogen diagnose, HRM has also been used for MAS to enhance the quality of soybean [113], rice [114], strawberry [115], and barley [116].

These methods normally do not reach the same high-throughput format than microarrays (Figure 1). However, nowadays, there are TaqMan and KASP arrays which help to process a high sample throughput for mid-density genotyping. In the case of TaqMan SNP genotyping, there are pre-designed and custom assays. Regarding HRM, as of recent, there are no commercial panels; however, it is the user who is in charge of designing and carrying out the assays. In the case of HRM and TaqMan (but not KASP) analyses, a low degree of multiplexing is possible (i.e., duplex).

2.1.3. Mass Spectrometry SNP Genotyping

Primers are designed immediately adjacent to the SNP locus and an SBE is carried out using mass-modified dideoxynucleotide terminators. The mass of the allele-specific product is determined by using matrix-assisted laser desorption–ionization time-of-flight (MALDI-TOF) mass spectrometry. Like all the other SNP genotyping technologies, this is used with identification purposes in crops. Besides, it is applied in MAS for quality traits in cereals, such as barley [117], rice [118], legumes (including pea) [119], and mung bean [120].

This is a high-throughput technology (Figure 1) which can process thousands of samples per day, which also allows the simultaneous amplification and detection of multiple markers per reaction (i.e., Agena iPLEX Gold, previously known as Sequenom iPLEX Gold). This method avoids the problems derived from a background signal typical from those based on hybridization. As the previous ones, this type of assay can be custom-designed.

2.1.4. SNP Analysis by NGS

With the increasing affordability of sequencing methods, these SNP genotyping platforms based on sequencing are becoming very popular. The main strategy nowadays consist of building reduced representation libraries (RRLs). By reducing the complexity of the targeted genome (normally digesting it with restriction enzymes), the depth of the sequencing can be increased. Among all the developed methods, including restriction site-associated DNA sequencing (RAD-Seq), diversity array technology sequencing (DArT-Seq), restriction fragment sequencing (REST-Seq), multiplex shotgun genotyping (MSG), sequence-based genotyping (SBG), specific-locus amplified fragment sequencing (SAF-Seq), etc., one of the most widely used in crops is genotyping by sequencing (GBS). Briefly, the whole genome is fragmented using restriction enzymes and short-read sequencing is performed on the ends (paired-end sequencing). Libraries for each sample are prepared using different barcodes; thus, a multiplex approach in which thousands of genotype SNPs across thousands of samples simultaneously was possible (Figure 1). GBS is used in studies on some traits that influence the nutritional value of food crops, such as the soluble solid content in plum [121]; sugar and acid content in apple [122]; sugar and carotenoid content in melon [123]; and certain mineral content in maize [124], pea [125], and spinach [126].

As in the previous technologies, pre-designed assays are available for some crops though custom panels of markers are also possible.

Thanks to this profusion of technologies that are becoming more and more affordable, a large number of SNP databases in crops is made available (Table 1). The data that have been made public in this way feed back into the agrigenomic field, as they can be used by other researchers to design their assays. Some of them only include marker information, but others also supplied the genotypes in different accessions (cultivars and wild crops relatives) as well as other useful tools, including genetic maps, genome sequences, etc. Table 1 clearly shows a higher representation of staple crops (i.e., cereals), given the very intense genetic breeding in recent decades, though other crops with a great economic importance, such as fruit trees (i.e., within Rosaceae family) or vegetables (i.e., tomato), are also present.

Table 1. List of the main public SNP databases in food crops. The type of information available ranges from the marker description to the genotype and map and/or genome location.

Database Name	Url	Crop ‡
CerealsDB	https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php, accessed on 17 January 2022	Bread wheat (*Triticum eastivum* L.)
Chickpea SNP-InDel Database (CicArVarDB)	https://cegresources.icrisat.org/cicarvardb, accessed on the 17 of January 2022	Chickpea (*Cicer arietinum* L.)
CropSNPdb	http://snpdb.appliedbioinformatics.com.au/, accessed on 17 January 2022	Bread wheat (*T. eastivum* L.) Cabbage (*Brassica rapa* L.) Cauliflower (*Brassica oleracea* L.) Indian mustard (*Brassica juncea* L.) Oilseed rape (*Brassica napus* L.)
Cucurbit Genomics Database (CuGeDG)	http://cucurbitgenomics.org/, accessed on 17 January 2022	Cucumber (*Cucumis sativus* L.) Melon (*Cucumis melo* L.) Pumpkin (*Cucurbita* spp.) Watermelon (*Citrullus lanatus* Thumb.)
Genome Database for Rosaceae (GDR)	https://www.rosaceae.org, accessed on 17 January 2022	Apple tree (*Malus* spp.) Blackberry (*Rubus* spp.) Peach tree (*Prunus* spp.) Pear tree (*Pyrus* spp.) Strawberry (*Fragaria* spp.)
Gramene	https://www.gramene.org, accessed on 17 January 2022	African rice (*Oryza galberrina* Steud) Asian rice (*Oryza sativa* L.) Barley (*Hordeum vulgare* L.) Foxtail millet (*Setaria italica* (L.) Beauv.) Maize (*Zea mays* L.) Sorghum (*Sorghum bicolor* (L.) Moench) Wheat (*Triticum* spp.)
Kazusa Tomato Genomics Database (KaTomicsDB)	https://www.kazusa.or.jp/tomato/, accessed on 17 January 2022	Tomato (*Solanum lycopersicum* L.)
Lettuce Genome Database (LettuceGDB)	https://www.lettucegdb.com, accessed on 17 January 2022	Lettuce (*Lactuca sativa* L.)
Maize Genetics and Genomics Database (MaizeGDB)	https://www.maizegdb.org/, accessed on 17 January 2022	Maize (*Z. mays* L.)
Maize SNP-DNA Fingerprint Database	http://doi.org/10.3390/agriculture11070597 (Tables S1 and S2; [127]), accessed on 18 January 2022	Maize (*Z. mays* L.)
Q-TARO (QTL Annotation Rice Online) database	http://qtaro.abr.affrc.go.jp/index.html, accessed on 18 January 2022	Asian rice (*O. sativa* L.)
SNP genotype database for avocado	https://doi.org/10.1007/s11295-019-1374-1 (Table S2; [128]), accessed on 18 January 2022	Avocado (*Persea americana* Mill.)
Sol Genomics Network	https://solgenomics.net, accessed on 18 January 2022	Tomato (*S. lycopersicum* L.)
SorGSD	https://ngdc.cncb.ac.cn/sorgsd, accessed on 18 January 2022	Sorghum (*S. bicolor* (L.) Moench)
SpinachBase	http://www.spinachbase.org, accessed on 19 January 2022	Spinach (*Spinacia oleracea* L.)
Rice SNP-Seek Database	https://snp-seek.irri.org, accessed on 19 January 2022	Asian rice (*O. sativa* L.)

Table 1. *Cont.*

Database Name	Url	Crop ‡
The IPK Crop EST Database (CR-EST)	http://pgrc.ipk-gatersleben.de/cr-est, accessed on 19 January 2022	Barley (*H. vulgare* L.) Bread wheat (*T. eastivum* L.) Pea (*Pisum sativum* L.) Potato (*Solanum tuberosum* L.)
The Tomato Integrated Database (Tomatonics)	http://plantomics.mind.meiji.ac.jp/tomatomics, accessed on 19 January 2022	Tomato (*S. lycopersicum* L.)
TropGENE-DB	http://tropgenedb.cirad.fr/tropgene/JSP/index.jsp, accessed on 19 January 2022	Asian rice (*O. sativa* L.) Banana (*Musa acuminata* Juss.) Bread fruit (*Artocarpus altilis* (Parkinson) Fosberg) Cassava (*Manihot esculenta* Crantz) Clemetine (*Citrus clementina* L.) Cocoa (*Theobroma cacao* L.) Coconut (*Cocos nucifera* L.) Coffee (*Coffea canephora* L.) Cupuassu (*Theobroma grandiflorum* Schum.) Oil palm (*Elaeis guineensis* Jacq.) Pummelo (*Citrus grandis* (L.) Osbeck) Sorghum (*S. bicolor* L. Moench) Sugarcane (*Saccharum officinarus* L.) Sweet orange (*Citrus sinensis* Osbeck)
Vitis International Variety Catalogue (VIVC)	https://www.vivc.de/index.php?r=site%2Findex, accessed on 19 January 2022	Grapevine (*Vitis* spp.)

‡ Even if there are more species in some databases, they were not included if there is no SNP information available or they are not food crops.

For the above, SNPs are the preferred markers to both, carry out genetic studies and undertake breeding programs in crops. Actually, genotyping assays have been developed for a large number of plants, including all major crops.

2.2. Nutritional and Phytochemical Profiles Assessed by Analytical Methods

The "omic" era has also reached the characterization of food plants in terms of their nutritional content, making use of metabolomic technologies. Thus, it is now possible (though still prohibitive, in many cases) to obtain the complete profiles of phytochemicals in complex extracts in a high number of samples. In this way, the compounds are identified by metabolic profiling and then quantified by target analysis. This has huge potential in plant breeding, especially in crop biofortification, which is still to be fully exploited. The different techniques normally used for metabolome analysis are enlisted here very briefly, as that is not the main scope of this review.

2.2.1. Mass Spectrometry (MS)

This is a very sensitive analytical technique, either used directly (non-hyphenated methods) or coupled with others (hyphenated methods), such as gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE). In the first case, it is possible to process a high number of samples in a short period of time, though the identification capacity is limited. The hyphenated methods, on the other hand, are undoubtedly more powerful when it comes to identifying and quantifying metabolites, and there is also the possibility of reducing the running times by using more advanced techniques in chromatography (i.e., ultra-high-performance liquid chromatography (UPLC) instead of high-performance liquid chromatography (HPLC) [129]). In any case, the metabolite identification generally requires the availability of libraries in order to compare the spectra obtained.

2.2.2. Nuclear Magnetic Resonance (NMR)

This is a very reproducible spectroscopic technique used to quantify metabolite levels. It allows a high-throughput process of samples, though it is generally less sensitive and has less resolution power than MS. Moreover, it is a non-destructive method, which makes it the perfect choice for studying the metabolome evolution (for instance, in different plant stages), instead of simply obtaining a snapshot of the plants at a particular moment.

Both techniques can be actually combined, resulting in the detection of a higher number of metabolites.

Until recently, the most common nutritional studies in food crops have focused on the quantification of a discrete number of compounds with a high impact in their nutritional value (targeted metabolic studies), though some widely targeted metabolomics analyses are starting to be carried out even in minor crops [130]. The initial steps which deal with the germplasm evaluation for nutrients and health-promoting compounds are essential for harnessing the biodiversity harbored by cultivated varieties, but also by breeding material and crop wild relatives. Some examples of these characterization works can be found in all groups of food crops, cereals [131], fruits [132], legumes [133], and vegetables [134,135], among others. In this sense, a considerable number of researches has compared different plant material within the same crop (for instance, landraces vs. commercial varieties) in order to identify outstanding accessions for future breeding programs aimed at enhancing the content of nutritious and beneficial compounds (reviewed in [3]). Metabolomic offers the opportunity to study the huge range of metabolites present in a sample (untargeted metabolic studies) and not only some specific compounds.

Another metabolome approach, apart from profiling commented above, consists of performing metabolomic fingerprints, where compounds are not individually identified. However, the metabolite profiles are compared among samples, for instance, to study the plants at different developmental stages [136] or under several biotic [137] and abiotic [138] stresses. We will not go into depth in the latter, as it is not related to the subject of this review, though it is noteworthy to mention that some studies use a combination of both approaches, i.e., by carrying out metabolomic fingerprint experiments in which the compounds are actually identified [136].

3. Association between the Traits of Interest and the Genomic Regions: Fishing for Genes

On one hand, one of the most useful and exploited genetic tools in crop breeding has been the linkage maps. Large SNP genotyping arrays have been used to build high and ultra-high-density genetic maps that allow the efficient marker-assisted selection (MAS) of beneficial alleles for the traits of interest. Nowadays, there are consensus and saturated genetic maps (mainly built with SSR and SNP markers) in virtually all the important crops and, in many cases, they are used to localize quantitatively trait loci (QTL). This fine mapping (often together with the QTL analysis) has led to the identification and cloning of the underlying gene(s), mainly in cereals (i.e., barley, maize, rice, and wheat), but also in some legumes (i.e., soybean) and vegetables (i.e., tomato) [139], though there are few cases for traits related to their nutritional value. An emerging application involves integrating metabolic/metabolomic and quantitative data to render metabolic QTL (mQTL). Until now, a number of these studies have been carried out, mainly in cereals (wheat, barley, rice, and maize) but also in oilseed rape and tomato [140]. As a result, numerous mQTL have been identified in those crops and some of them have eventually led to the identification of putative candidate genes controlling metabolic traits [140].

On the other hand, in genomics (the field that concerns us in this review), the whole genome of an organism is studied. As could be expected, the development of NGS technologies has led to a real boost for its applications, such as genome-wide association studies (GWAS). With the SNP genotyping by NGS, it is possible and affordable to rapidly scan markers across the complete genome of many individuals to find variations associated with a particular trait. In fact, the genotypes for thousands of SNPs are currently available for many crop species, as shown in Table 1. In order to make the

most of all this already existing information, it can be combined with the results derived from the technology to analyze metabolites. In this line, researches which combine metabolic/metabolome and genome association results (metabolic/metabolomic GWAS, mGWAS) are starting to be carried out in crops (Table 2) and they are expected to become very helpful in genomic-assisted breeding programs by whole-genome selection and eventually in identifying some of the genes potentially influencing the nutritional value and the content of health-promoting compounds.

A potential drawback of this methodology, especially in the case of complex traits (as is the case of metabolism-related traits), is that the most significant variant obtained (i.e., allele of a SNP) is sometimes not responsible for metabolic differences. Actually, it is also common, as in any statistical analysis, to obtain spurious associations, for instance, when the trait heritability is low (high environmental effect). For this reason, it will still be necessary to carry out the validation of the candidate genes identified. In this sense, in many of those mGWAS involving compounds with a potential use to biofortify the respective crop (Table 2), other "omics" technologies, mainly transcriptomics, have assisted researchers in untangling the relationships between genotype and phenotype and in pinpointing the causal gene(s). Furthermore, it is also common to validate those findings by using mutants (knockout and/or overexpressing lines) and transgenic plants. Such an encompassing approach will undoubtedly speed up the process of obtaining healthier and nutritionally richer crops. Even if it is not the purpose of many of those studies, aimed at evaluating the metabolic changes that plants undergo during their development or to face environmental challenges (i.e., biotic and abiotic stresses), that knowledge about the genes responsible for the changes in metabolite contents is applicable in order to enhance the food in phytochemicals with beneficial properties.

Table 2. Metabolomic genome-wide association studies (mGWAS). Only groups of compounds that play an important role in human nutrition and/or health status are shown.

Crop	Species	Analytical Technique ‡	Metabolite	Reference
Apple tree	*Malus* × *domestica* Borkh.	UHPLC–ESI-QTOF-MS, NMR	Flavonoids, polyphenols, sugars, terpenoids	[141]
Barley	*H. vulgare*	HPLC-FL, HPLC-MS, IC-MS/MS	Amino acids, glutathione, organic acids, starch, sugars, vitamin E (tocopherol)	[142]
		HPAEC-PAD, HPLC-ELSD, HPLC-MALDITOF-MS	Sugars	[143]
		HPLC-Fluorescence detection	Carotenoids (i.e., tocopherols and tocotrienols: vitamin E)	[144]
Barley Bread wheat Maize Potato Rice Sweet orange tree	*H. vulgare* *T. aestivum* *Z. mays* *S. tuberosum* *O. sativa* *Citrus x sinensis* (L.) Osbeck	GC-TOF-MS	Flavonoids	[145]
Blueberry	*Vaccinium* spp.	GC-MS	Fatty acids, phenylpropanoids, terpenoids	[146]
Bread wheat	*T. aestivum*	GC-MS	Amino acids, organic acid ‡‡ sugars	[147]
Foxtail millet	*S. italica*	HPLC-ESI-QTRAP-MS/MS	Alkaloids, amino acids, fatty acids, organic acids, phenolamides, polyphenols (i.e., flavonoids, anthocyanins...), sugars, vitamins	[148]

Table 2. *Cont.*

Crop	Species	Analytical Technique ‡	Metabolite	Reference
Lettuce	*L. sativa*	GC-TOF-MS	Alkaloids, amino acids, organic acids, polyamines, polyphenols, sugars, vitamins, etc.	[149]
Loquat	*Eriobotrya japonica* (Thunb.) Lindl.	UPLC-ESI-MS/MS	Alkaloids, flavonoids, phenolic acids, polysaccharides, terpenoids	[150]
Maize	*Z. mays*	LC-MS/MS	Fatty acids	[151]
		LC-ESI-(QTRAP or QqTOF)-MS/MS	Amino acids, fatty acids, flavonoids	[152]
		GC-MS	Amino acids, organic acids, phenylpropanoids	[153]
		HPLC-Fluorescence detection	Tocochromanols (tocopherols and tocotrienols)	[154]
		HPLC-PDA	Carotenoids	[155]
		UPLC-HRMS	Amino acids, fatty acids, flavonoids, benzoxazinoids, terpenoids	[156]
		HPLC, UPLC	Carotenoids	[157]
		CEC	Amino acids	[158]
		LC-ESI-QqTOF-MS/MS	Flavonoids	[159]
		HPLC	Carotenoids	[160]
		UPLC-PDA	Tocopherol (part of vitamin E)	[161]
		GC-TOF-MS	Amino acids, (poly)amines, organic acids, sugars, vitamin E (tocopherol)	[162]
		HPLC-PDA, HPLC-fluorescence detection	Carotenoids, phenolics, tocopherol (a form of vitamin E)	[163]
		HPLC-fluorescence detection	Carotenoids (i.e., tocopherols and tocotrienols: vitamin E)	[164]
		HPLC-PDA	Carotenoids (α-carotene, β-carotene, β-cryptoxanthin, lutein, phytofluene, zeaxanthin, zeinoxanthin)	[165]
		HPLC-UV/Vis	Anthocyanins	[166]
Potato	*S. tuberosum*	UPLC-Q-TOF-MS	Alkaloids, amino acids	[167]
Rice	*O. sativa*	LC-ESI-Q TRAP-MS/MS	Phenolamides	[168]
		GC-TOF-MS	Amino acids, flavonoids, organic acids	[169]
		LC-ESI-MS/MS	Amino acids, fatty acids, flavonoids	[170]
		HPLC-ESI-QTOF/MS	Amino acids, flavonoids, phenolamines, terpenoids	[171]
		HPLC-ESI-(QTRAP or QqTOF)-MS	Amino acids, flavonoids, phenolamines, terpenoids	[70]
		LC-ESI-Q TRAP-MS/MS	Flavonoids	[172]

Table 2. *Cont.*

Crop	Species	Analytical Technique ‡	Metabolite	Reference
Soybean	*Glycine max* L.	GC	Fatty acids	[173]
		HPLC-DAD	Isoflavones	[174]
		HPLC-MS	Aminoacids, isoflavones, lipids, organic acids	[175]
Tea	*Camellia sinensis* L.	HPLC	Theanine, caffeine, catechins	[176]
		HPLC-PDA	Amino acids, caffeine, catechins	[177]
Tomato	*S. lycopersicum*	GC-MS	Organic acids, sugars	[178]
		GC-MS	Amino acids, organic acid ‡‡, sugars	[179]
		HPLC-MS/MS	Alkaloids ‡‡‡	[180]
		GC-MS	Fatty acids, lipids, carotenoids (i.e., tocopherols and tocotrienols: vitamin E)	[181]
Wheat	*T. aestivum*	HPLC-ESI-QTRAP-MS/MS	Amino acids, (poly)amines, flavonoids, organic acids, sugars, vitamins, etc.	[182]

‡ CEC: cation exchange chromatography; ELSD: evaporative light scattering detection; GC: gas chromatography; GC-MS: GC mass spectrometry; GC-TOF-MS: GC time-of-flight mass spectrometry; HPAEC-PAD: high-pH anion-exchange chromatography with pulsed amperometric detection; HPLC: high-performance liquid chromatography; HPLC-ESI-(QTRAP or QqTOF)-MS: HPLC-ESI-quadrupole TRAP or TOF tandem mass spectrometry; HPLC-MALDITOF-MS: HPLC matrix-assisted laser desorption–ionization time-of-flight mass spectrometry; IC-MS/MS: ion chromatography tandem mass spectrometry; LC-ESI-MS/MS: liquid chromatography–electrospray ionization tandem mass spectrometry; LC-Q-TOF-MS: liquid chromatography quadrupole TOF mass spectrometry; NMR: nuclear magnetic resonance; UPLC-ESI-MS/MS: ultra-high-performance liquid chromatography ESI tandem mass spectrometry; UPLC-HRMS: UPLC high-resolution mass spectrometry. ‡‡ Oxalic acid (anti-nutrient). ‡‡‡ Steroidal glycoalkaloids (SGAs): most of them are considered anti-nutrients.

4. Introducing Allelic Variants to Biofortify Crops

The last stages of the biofortification process in crop plants can be tackled through different approaches, including both conventional and modern biotechnology techniques, such as transgenesis, cisgenesis, intragenesis, or gene editing (i.e., CRISPR/Cas), in order to introduce genetic variation into the gene pool of the crop. Here, we will describe conventional breeding, transgenesis, cisgenesis, and intragenesis, as well as their applications in crop biofortification.

4.1. Conventional Breeding Assisted by Genomic Tools

Biofortification through conventional breeding is based on crosses within a sexually compatible group, specifically between donor plants with nutritional properties of interest and recipient ones with good agronomic characteristics. Many types of populations have been developed to perform genetic mapping, QTL identification, and association studies (i.e., both temporal (F_2, backcrosses (BCs) and advance backcrosses (ABs)) and immortal (double haploid lines (DHLs), recombinant inbred lines (RILs), near isogenic lines (NILs), multi-parent advanced generation inter-cross (MAGIC), and nested association mapping (NAM)) ones). Among them, the most widely used in plant breeding to introgress DNA regions that harbor beneficial alleles for the trait of interest from the donor into the recipient parent are ABs, NILs, and RILs (Figure 2).

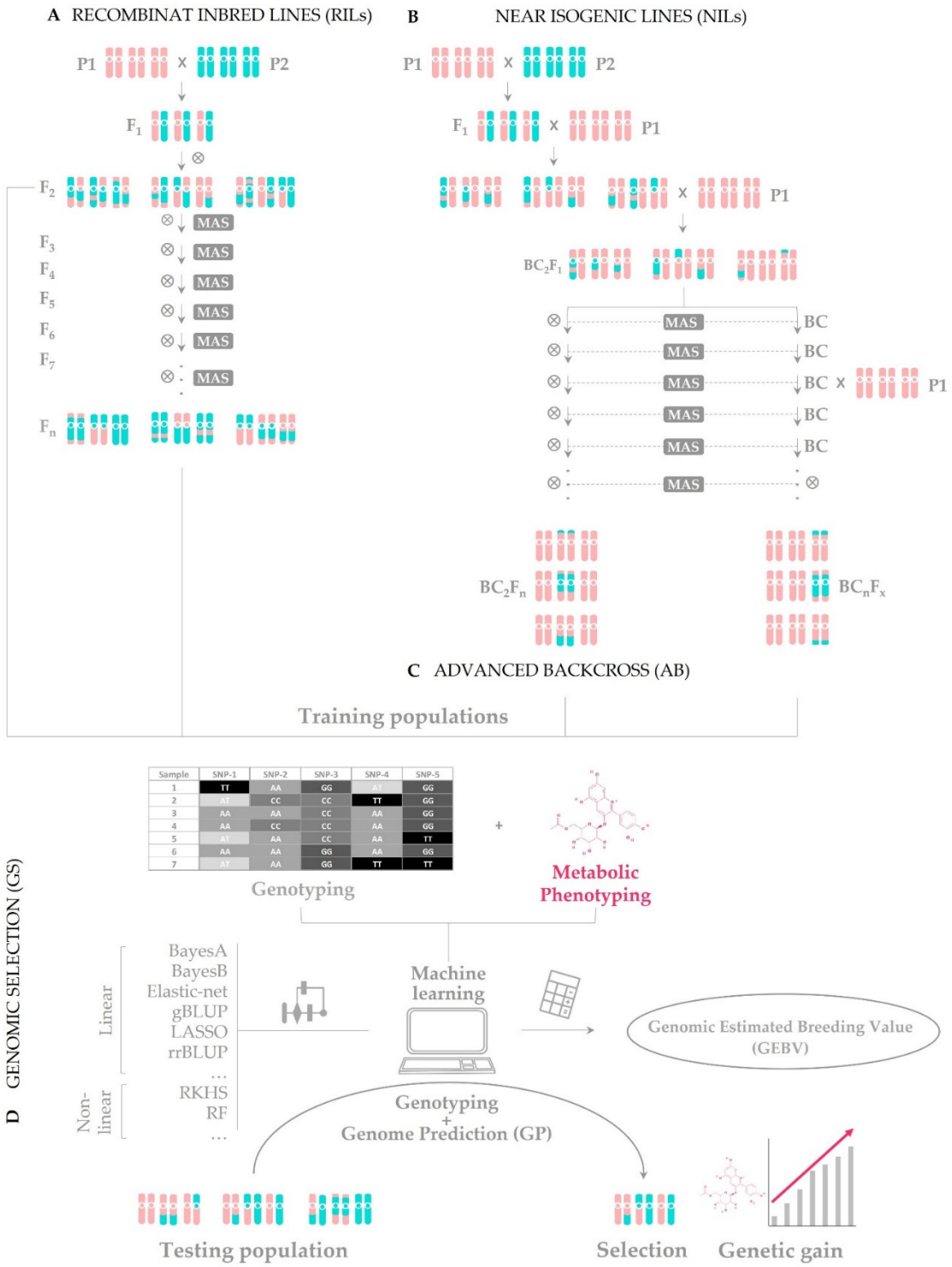


Figure 2. Construction of breeding populations: (**A**) recombinant inbred lines (RILs); (**B**) near isogenic lines (NILs); (**C**) advanced backcross; and (**D**) their use for Genomic Selection (GS). Only some of the possible crossing designs are shown.

Several generations are needed, so it takes a substantial amount of time to obtain crops with the desired nutritional and agronomic characteristics by using this strategy. An emerging alternative approach to save time, effort, and money consists of carrying out a genomic selection (GS), as coined by Meuwissen et al. [183], based on a genomic prediction (GP) (Figure 1D). Instead of phenotyping at every stage of the population building (like

in the MAS strategy), it is only carried out in what is known as training population (DHs, F_2, marker-assisted recurrent selections, etc.). These data, together with genome-wide genotypic data from that same training population, are used to calculate the genomic estimated breeding value (GEBV) through processes of machine learning by means of different regression models. So, GEBV is a parameter used to quantify the genetic merit of a certain individual in order to improve the crop in the trait of interest. Finally, the GP is carried out with the data coming from genotyping the testing population (the breeding population) without the need to phenotype it. In this way, the individuals selected by the testing population are expected to show a genetic gain, i.e., an increase in performance thanks to the gene variant(s) responsible for the aforementioned trait. With this method, all markers are taken into account, not only those which show a significant association with the trait (as in MAS); thus, loci with little additive effects can also be detected. Until now, this approach has been scarcely used in crops for metabolite and nutritional content, such as in tomato [184] and wheat [185].

Furthermore, the chances of achieving biofortification by conventional breeding depends on the crop itself, since the strategy relies on the genetic variability available within its gene pool, which is usually limited in commercial varieties. This could be overcome by crossing plants with landraces or with more distant wild relatives that normally harbor higher genetic variability and, sometimes, can be richer in nutrients [3,186]. However, in some cases, it would be impracticable to obtain biofortified crops using conventional breeding. That would be the case when the genetic variability needed for a specific trait is insufficient within the gene pool, or when the investment of time and resources would be excessive, especially with non-diploid species, when the trait heritability is low or when linkage drag is unavoidable.

In spite of its limitations, conventional breeding is currently the most accepted method, as it is sustainable and it is not subject to regulatory obstacles. Nowadays, an important number of crops have been conventionally bred to enhance their nutritional content. In fact, several international organizations have initiated different programs to accomplish this objective. Harvest Plus, launched in 2003, is the most important one and is focused on enhancing the content of provitamin A, iron, and zinc in staple food crops across Asia and Africa [187]. It has managed to biofortify a large number of crops, many of which have been already released. Until 2019, there is a total of 242 across 30 developing countries [188]. Different studies have demonstrated the efficacy of biofortification through conventional methods, specifically increasing the content of micronutrients [189,190]. Furthermore, other smaller institutions are working on developing conventionally biofortified crops. For example, the International Potato Centre (CIP) has obtained, tested, and advertised an orange sweet potato enriched in provitamin A [191], and the International Maize and Wheat Improvement Centre (CIMMYT) has released different hybrid varieties with increased levels of the amino acids lysine and tryptophan through the incorporation of the naturally occurring mutation *opaque-2* (*o2*) into different maize varieties [192].

The assistance of genomic tools has facilitated the development of many conventional biofortified crops, as they allow breeders to exploit the available genetic variability more efficiently; thus, time and costs can be significantly reduced. Plant breeding has existed since plant domestication started around 10,000 years ago, and the selection carried out at the beginning merely attends to the phenotype. However, with the application of genetic and genomic tools, genetic variants can be associated with differences in phenotypes, which then enables the selection at early stages of the plant. For that, the construction of genetic maps has been essential, as previously mentioned. Many studies have found markers linked to genes or QTL which can control the content of nutritional compounds, for example, those related with carotenoid variation in sorghum [193], mineral micronutrients in beans and wheat [194,195], vitamins levels in different cereal crops [196], etc. Thus, individuals with the best gene combination have been identified and used as potential donors in breeding programs to enhance the content in micronutrients (minerals and

vitamins) and health-promoting compounds (polyphenols, carotenoids) in all kinds of crops, including cereals, fruits, legumes, and vegetables (Table 3).

Table 3. Biofortified crops through different techniques.

Technique	Crop	Method	Biofortified Trait	Reference
Conventional breeding	Rice	Backcrosses between a high-yielding cultivar and the IR68144 line	A 2.54-fold increase in iron and 1.54-fold increase in zinc	[4]
	Maize	Backcrosses involving diverse exotic donor lines	Lines with high provitamin A content by accumulating mainly high β-carotene and lines with high provitamin A by promoting accumulation of high levels of both carotenes and xanthophylls	[13]
		Marker-assisted introgression of *lpa1-1* and *lpa2-1* alleles in elite lines of provitamin A-enriched quality protein maize (QPM)	A reduction in phytic acid content and improvement in the mineral bioavailability in lines of QPM rich in provitamin A	[197]
		Introgression of *VTE4* (γ-tocopherol methyl transferase) allele into four provitamin-A rich QPM elite inbreds using marker-assisted backcross breeding	An increase in α-tocopherol to 15.2 ppm over 8.0 ppm in the original inbreds	[14]
	Wheat	Marker-assisted introgression of group 4 and 7 chromosomes of the wild ancestor *Aegilops peregrina* in a commercial variety of wheat	Higher content in iron and zinc in wheat grains	[5]
		Backcrosses between low-yielding exotic donor lines and commercial varieties	Black, purple, and blue lines with high content in anthocyanins	[38]
	Cassava	Rapid cycling recurrent selection	Significant gains for total carotenoid content and total β-carotene	[15]
	Potato	'Atlantic' and 17 4x-2x hybrids between *S. tuberosum* and diploid hybrids of *Solanum phureja-Solanum stenotomum*	Higher contents of copper, iron, manganese, and zinc	[6]
	Tomato	Backcrosses between landraces of tomato	Hybrid with increased concentration of polyphenols and high antioxidant activity in pink ripeness stage	[39]
	Bean	Backcrosses between low and high mineral genotypes using a QTL mapping approach	Increased iron and zinc content	[7]
	Chickpea	Crosses between different cultivars	Higher content of carotenoids	[16]

Table 3. *Cont.*

Technique	Crop	Method	Biofortified Trait	Reference
Transgenesis	Rice	Endosperm-specific overexpression of *Arabidopsis thaliana GTP cyclohydrolase I* (*GTPCHI*) and *aminodeoxychorismate synthase* (*ADCS*) genes	An enhancement of 100 times in folate	[198]
		Overexpression of *phytoene synthase*	Higher content in β-carotene	[17]
		Expression of four synthetic genes: *sZmPSY1*, *sPaCrtI*, *sCrBKT*, and *sHpBHY* (for phytoene synthase, phytoene desaturase, β-carotene ketolase, and β-carotene hydroxylase, respectively)	Synthesis de novo of the carotenoid astaxanthin	[40]
		Coexpression of an *Arabidopsis* nicotianamine synthase (*AtNAS1*), bean ferritin (*PvFerritin*), bacterial carotene desaturase (*CRTI*), and maize phytoene synthase (*ZmPSY*)	Simultaneous increase in iron, zinc, and β-carotene content in the rice endosperm	[29]
		Constitutive overexpression of the rice *GDP-L-galactose phosphorylase* (35S-*OsGGP*) gene	Increase in ascorbate concentrations in germinated brown rice	[18]
		Expression bacterial *aspartate kinase* (*AK*) and *dihydrodipicolinate synthase* (*DHPS*), downregulation of rice *lysine ketoglutarate reductase/saccharopine dehydrogenase (LKR/SD)* and selection of marker-free transgenic lines	Up to 25-fold increase in free lysine levels	[36]
		Expression of an AmA1 gene from *Amaranthus hypochondriacus*	A significant increase in the content of several EAAs, including lysine, threonine, and valine, as well as a 1.06~12.87% increase in the total protein content	[37]
	Maize	Overexpression of the bacterial genes *crtB* (for phytoene synthase) and *crtI* (for the four desaturation steps of the carotenoid pathway) under the control of a endosperm-specific promoter	An increase in total carotenoids of up to 34-fold with a preferential accumulation of β-carotene in the maize endosperm	[19]
		Endosperm-specific overexpression of soybean ferritin	A 2-fold improvement in seed iron bioavailability	[8]
		Coexpression of *Gm8gGCHI* and *GmADCS* genes driven by endosperm-specific promoters	A 4.2-fold increase in folate (vitamin B9) level in transgenic maize grains	[20]
		Insertion of the lysine-rich *sb401* gene	Significantly higher levels of lysine total protein in maize seeds	[35]
	Wheat	Constitutive expression of the rice *nicotianamine synthase 2* (*OsNAS2*) gene	Higher concentrations of grain iron and zinc, and enhanced localization of iron and zinc in endosperm and crease tissues, respectively	[9]
	Cassava	Coexpression of *ferritin* (*FER1*) and mutated *Iron transporter* (*IRT1*) from *A. thaliana*	Accumulation of iron levels 7–18 times higher and zinc levels 3–10 times higher	[10]

Table 3. *Cont.*

Technique	Crop	Method	Biofortified Trait	Reference
	Potato	Overexpression of *AtGTPCHI*, *AtADCS*, *OsHPPK/DHPS* and *AtFPGS* genes	A 2-fold increase in folate content in mature tubers and stable accumulation of folates for up to 9 months of storage	[21]
		Simultaneous expression of *Wrinkled 1* (*WRI1*), *Diacylglycerol acyltransferase 1* (*DGAT1*) and *Oleosin* under the transcriptional control of tuber-specific (patatin) and constitutive (CaMV-35S) promoters.	Over a 100-fold increase in triacylglycerol accumulation to levels up to 3.3% of tuber dry weigh	[33]
	Sweet Potato	Expression of a barley *NA synthase 1* (*HvNAS1*) gene	A 3- and 2.9-fold increase in the concentrations of iron and zinc, respectively	[11]
	Tomato	Cross between *GTPCHI* and *ADCS* overexpressing plants	A 25-fold more in folate (Vitamin B9) level in fruits	[199]
		Overexpression of an *A. thaliana Orange* (*AtOR*) gene	An increase in total carotenoids in fruits	[22]
		Overexpression of *GDP-l-galactose phosphorylase* (*GGP*) gene from *Actinidia chinensis* under the control of the 35S promoter	A 3- to 6-fold higher content in ascorbic acid in fruits	[200]
		Fruit-specific expression of the transcription factor *AtMYB12*	Increased content of different phenylpropanoids	[23]
	Strawberry	Overexpression of a *GDP-l-galactose phosphorylase* (*GGP*) gene from *Actinidia chinensis* under the control of the 35S promoter	A 2-fold higher content in ascorbic acid in fruits	[200]
	Banana	Expression of a *Fe'i banana-derived phytoene synthase* (*MtPsy2a*) gene under the maize polyubiquitin promoter	Enhanced β-carotene content in fruit	[24]
	Soybean	Overexpression of the bacterial genes *crtB* (for phytoene synthase) and *crtW* and *bkt1* (ketolase genes) under the control of seed-specific promoters	Enhanced accumulation of ketocarotenoids in seeds	[201]
		Overexpression of adenosine 5′-phosphosulfate sulfurylase 1	Higher amounts of sulfate, cysteine, and some sulfur-containing secondary metabolites in seeds	[34]
		Overexpression of a *GmDGAT2A* gene driven by a seed-specific promoter of *Gmole1*	Significantly increased linoleic acid content specifically and total oil content	[32]
	Bean	Seed-specific overexpression of a *GTP cyclohydrolase I* gene from *Arabidopsis* (*AtGchI*)	Increased folate levels in raw desiccated seeds by up to 3-fold	[25]
	Canola	Downregulation of lycopene ε-cyclase (ε-*CYC*)	Increased levels of β-carotene, zeaxanthin, violaxanthin, and lutein	[26]
	Brassica carinata	Expression of an *18-carbon ω3 desaturase* (*CpDesX*) gene from *Claviceps purpurea* and a *20-carbon ω3 desaturase* (*Pir-ω3*) gene from *Pythium irregulare*	Up to 25% increase in eicosapentaenoic acid	[31]
	Linseed	Expression of a Δ6-desaturase from *Primula vialii*	Transgenic lines that accumulate the omega-3 fatty acid stearidonic acid	[30]

Table 3. *Cont.*

Technique	Crop	Method	Biofortified Trait	Reference
Cisgenesis	Barley	Expression of a barley *phytase* gene (*HvPAPhy_a*)	Decrease in phytate concentration, which then increases phosphate bioavailability	[202]
	Potato	Suppression of a *starch phosphorylase* L. gene through dsRNAi technology	Decrease in starch degradation what reduces the accumulation of reducing (glucose, fructose) and non-reducing (sucrose) sugars in tubers stored at 4 °C	[203]
	Apple	Expression of *MdMYB10* transcription factor	Red-fleshed 'Gala' apples rich in anthocyanins	[41]
Intragenesis	Potato	Silencing of a *granule-bound starch synthase* (*GBSS*) gene	An increase in amylopectin content	[204]
		Silencing of an *asparagine synthase gene* (*StAs1*)	Reduced free asparagine concentration by up to 80% and consequent decrease in acrylamide content in processed potato	[205]
		Overexpression of a *lycopene b-cyclase* (*StLYCb*) gene under the GBSS promoter	An increase in β-carotene accumulation in potato tubers	[27]
	Tomato	Suppression of a *DE-ETIOLATED1* (*DET1*) gene through RNAi technology	Enhanced carotenoid and flavonoid content	[28]
	Wheat	Suppression of a *γ-gliadin* gene by using RNAi technology	Gluten-free wheat	[206]
		Overexpression of a *vacuolar Iron transporter* (*TaVIT2*) under the control of a wheat endosperm-specific promoter	An increase in more than 2-fold of iron in white flour fractions	[12]
	Soybean	RNAi technology	Plenish® high oleic	Dupont-Pioneer (Johnston, IA, USA)
			Vistive® Gold low saturated high oleic	Monsanto (St. Louis, MO, USA)

4.2. Modern Biotechnology Techniques

4.2.1. Transgenesis

In biofortification, transgenic approaches consist of the transference of one or more alleles from genes responsible for the increase in the nutritional value from one or more organisms to the crop of interest. They are really helpful in overcoming the main handicap of conventional breeding, i.e., the limited genetic variation within the same or sexually compatible species [207]. Moreover, genetic transformation through transgenesis can achieve the expression of a gene independently of its origin, in terms of evolution, taxonomy, and even kingdom [19,208,209]. Hence, when a specific nutrient or a bioactive compound is not naturally synthesized in a crop, transgenesis is the only way to engineer the crop to produce it. Therefore, this strategy helps to exploit a much larger gene pool and transfers more than one gene and their regulatory regions simultaneously (Figure 3A). In this way, the crop can be enriched in more than one nutrient at the same time, as it has already been successfully engineered in rice [29]. However, it is important to take into account that some crops are recalcitrant to transformation and/or regeneration, for example, some cereals [210] or legumes [211].

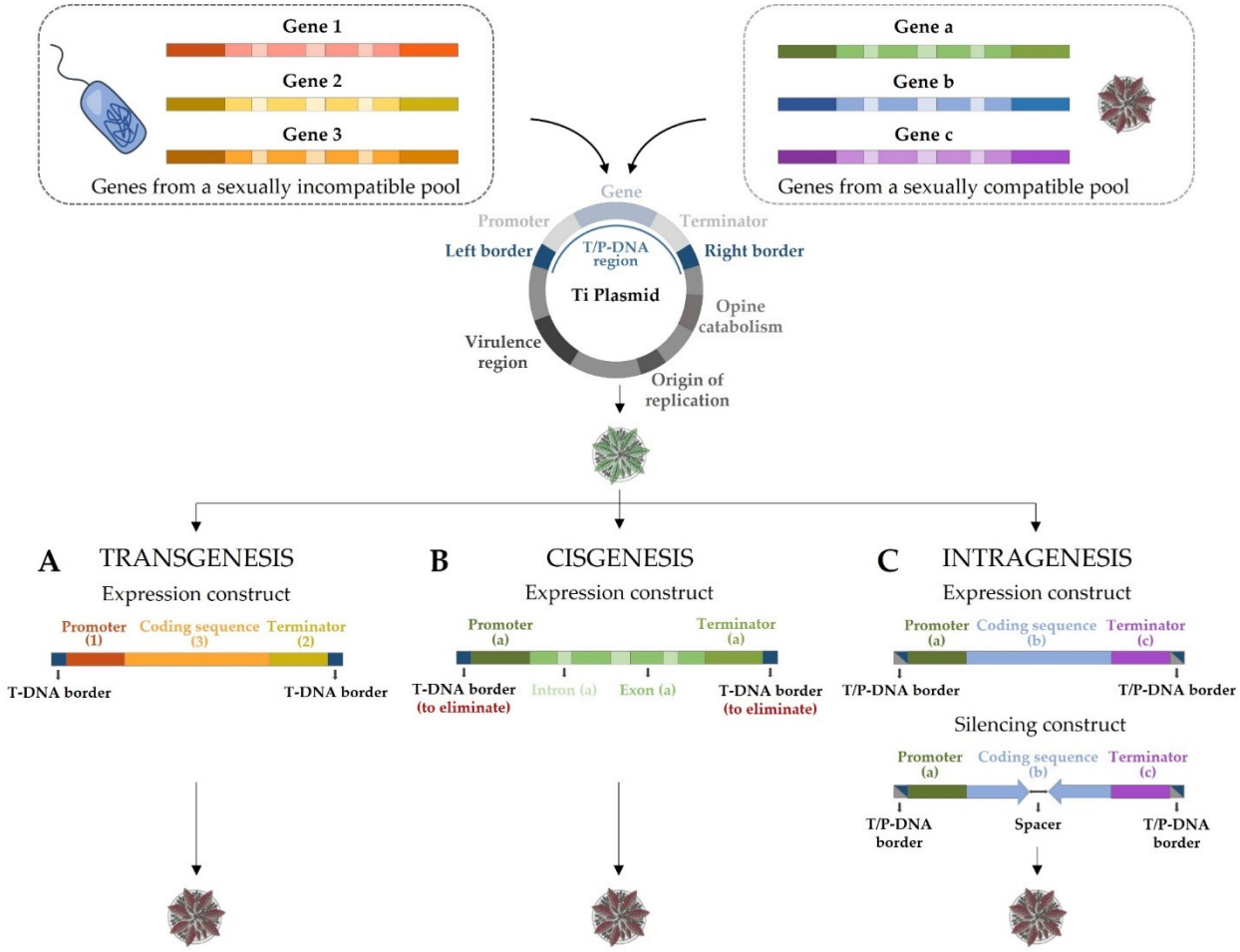


Figure 3. Schematic representation of three modern biotechnology techniques to introduce allelic variants of interest in a recipient organism: (**A**) transgeneis; (**B**) cisgenesis; and (**C**) intragenesis.

Transgenic approaches require a lot of time and resources. The identification and characterization of the gene(s) are needed to eventually introduce them in the crop. Nevertheless, transgenesis is less time-consuming than the conventional alternative and more cost-effective than the agronomic fortification, which is ineffective in the long term because it requieres regular application of fertilizers [212]. This, together with the absence of taxonomic constrictions and the possibility of designing almost any synthetic gene, has resulted in a big number of biofortified crops developed through transgenic strategies (Table 3). One of the most remarkable examples is Golden Rice, obtained to alleviate the vitamin A deficiency [17]. It was the first application of transgenic biofortification, in which a carotenoid-free rice endosperm was genetically engineered to produce β-carotene (provitamin A) by expressing the genes codifying for the phytoene synthase and the carotene desaturase [17]. In addition, a clinical trial in humans has demonstrated that Golden Rice could be an alternative source of vitamin A for adults [213]. As in the case of conventional breeding, many different strategies have been applied to almost any kind of crop, including cereals, legumes, vegetables, fruits, and oilseeds, whereby the targets of biofortification are fatty acids, essential amino acids, and antioxidants, among others (Table 3).

The main disadvantage of these crops is the strict regulation to which they are subject to, at least, in Europe (more deeply described further on). However, some biofortified crops have gone beyond this limitation and they have been released. Some of these crops are cassava with improved levels of zinc, iron, β-carotene, or proteins, released by Biocassava Plus; canola with a higher availability of phosphate due to phytate degradation,

released by BASF; and linseed enhanced in essential amino acids, released by the University Saskatchewan (Saskatoon, Canada).

4.2.2. Cisgenesis and Intragenesis

Cisgenesis and intragenesis are approaches that, to some extent, were developed to overcome the main limitation of transgenesis—its strict regulation [214]. The gene pool exploited here can only come from naturally crossable species; therefore, they might be a suitable alternative to obtain biofortified crops.

On the one hand, the terms "cisgenic plant" were first introduced in 2006 as "a crop plant that has been genetically modified with one or more genes (containing introns and flanking regions such as native promoter and terminator regions in a sense orientation) isolated from a crossable donor plant" [215] (Figure 3B). This donor plant has to belong to the same species than the modified crop or to a sexually compatible species; thus, the gene pool available for cisgenesis is identical to the gene pool exploited by conventional breeding. Nevertheless, unlike conventional breeding, only the gene(s) of interest, and no undesired sequences (linkage drag), are transferred to the final cisgenic crops.

On the other hand, the terms "intragenic plant" were introduced in 2004 and they refer to the isolation of specific genetic elements from a plant, the recombination of these elements in vitro, and the insertion of the resulting expression cassettes into a sexually compatible plant [216] (Figure 3C). Intragenesis can also be carried out using constructs with RNA interference (RNAi) [28,206] or genes edited, for instance, by CRISPR/Cas, as this technology has been successfully used to edit the genome of crops [217,218]. Therefore, intragenesis provides the possibility of creating novel combinations that render higher variability and novel expression patterns to develop new genetically modified organisms (GMOs) with new properties that will not happen spontaneously in nature or through conventional breeding.

The main difference between cisgenesis and intragenesis is related to the regulatory regions. In cisgenesis, the transgene is a complete DNA copy of the gene as it can be found in the donor plant (with promoter, introns, and terminator) in the normal-sense orientation (Figure 3B). In intragenesis, there is not any requisite about these regulatory elements, as long as all the genetic elements come from crossable donor plants, so that they can be engineered before being used in the transformation (Figure 3C). Consequently, intragenesis is not considered as close to conventional breeding as cisgenesis.

In both cases, when *Agrobacterium*-mediated transformation is used, T-DNA borders (flanking sequences of the DNA to be transferred) can be also inserted in the plant genome. This is a controversial topic as some authors are in favor of using T-DNA borders, claiming that they are safe because they are short non-coding sequences that can be found in plant genomes naturally too [219]. The evident argument against T-DNA borders is that all DNA sequences integrated into the recipient plant should come from a sexually compatible DNA pool, as established by both cisgenesis and intragenesis definitions [215,216]. Thus, both cisgenic and intragenic crops should be free of those T-DNA borders, and also of selection markers and vector backbones, as both of them are supposed to be genetically modified plants that do not contain foreign genes (only genes coming from cross compatible species). Two alternative solutions have been proposed. First, plants without T-DNA borders can be selected just by carrying out a PCR. In fact, the integration rate of the T-DNA borders in the plant genome is relatively low, as is the case of transgenic potatoes carrying R genes for late blight, in which only 45% of transformants possessed T-DNA borders [220]. Second, T-DNA border-like sequences found in the plant genomes, known as P-DNA borders, can be used upstream and downstream the gene to be transferred [216,221]. A rearrangement of the original gene is thus required, as it was in the donor plant, which is why this option should only be chosen in the case of intragenic plants. Furthermore, the presence of T-DNA borders in both types of plants could be a problem for the public acceptance and in terms of regulation [222]. Regarding the other non-plant sequences, the use of selection markers is not necessary when the transformation efficiency is high [223] or the product codified

by the introduced gene can be visually detected, including a pigmented compound (i.e., carotenes, anthocyanins) [224]. There are also methods to eliminate markers based on site-specific recombination (marker genes are flanked by specific recombination sites) [225], or by carrying out a co-transformation, which allows the segregation of the transgene and the marker gene in the progeny, as they are integrated in different positions of the genome [226].

In comparison to transgenesis, cisgenesis and intragenesis have two clear limitations (Table 4). The first one is that the available variability only exists in plants from the same sexual compatibility group, as in conventional breeding. However, this disadvantage could be overcome, to some extent, by gene edition (in the case of intragenesis) or by making use of the higher biodiversity present in landraces [3] or wild relatives [186]. The second limitation is the need to remove the selection markers and the vector backbones, which could be both time- and labor-consuming. On the other hand, although the three technologies are subject to the same regulation, cisgenic and intragenic crops are more accepted by the general public than transgenic ones [227–229].

Table 4. Comparison of the main characteristics of conventional breeding, transgenesis, cisgenesis, and intragenesis.

Characteristic	Conventional Breeding	Transgenesis	Cisgenesis	Intragenesis
Variability source	Sexually compatible group	Any organism	Sexually compatible group	Sexually compatible group
Method	Crosses and selection	Recombinant DNA	By *Agrobacterium*	By *Agrobacterium* (recombinant DNA)
Introducing DNA	Natural	Natural and/or artificial	Natural	Natural and/or artificial
Gene pool	Unaltered	Altered	Unaltered	Altered
Borders	-	T-DNA	T-DNA (to be eliminated)	T-DNA or P-DNA
Linkage drag	Yes	No	No	No
Expression modulation	No	Yes	Yes	Yes
Time	High	Medium	Medium	Medium

When compared to conventional breeding, cisgenesis and intragenesis are considered fast alternatives to transfer genes between plants from the same sexual compatibility group, especially for species with long lifetimes and high heterozygosity levels (Table 4). Additionally, these two approaches are able to avoid linkage drag issues associated with backcrosses in conventional breeding, as only the sequences of interest are transferred (Table 4). Changes in the gene expression levels can also be achieved with both techniques (Table 4). The introduction of the complete natural gene (cisgenesis) and changes in promoters and terminators (intragenesis) may increase the levels of expression, whereas the use of silencing constructs (intragenesis) could reduce them. Moreover, new genetic variability can be generated with different combinations of genetic elements with intragenic approaches.

Although most of the new traits incorporated to relevant crops through cisgenesis and intragenesis are related to disease resistance [216,225] and abiotic stress tolerance [230], these strategies have been also applied with biofortification purposes (Table 3). For example, Holme et al. [202] obtained a cisgenic barley by inserting copies of a barley phytase gene (*HvPAPhy_a*). Those barley plants with a single copy of the gene showed a 2.8-fold increase in the phytase activity and an enhanced bioavailability of phosphate. A cisgenic potato was developed by suppressing the *starch phosphorylase L* gene through dsRNAi (double-strand RNA interference) technology to decrease starch degradation [203]. Then, the accumulation of reducing (glucose, fructose) and non-reducing (sucrose) sugars was lower in tubers

stored at 4 °C. Finally, cisgenic red-fleshed apples, rich in anthocyanins, were developed by expressing the *MdMYB10* gene, a transcription factor involved in anthocyanin biosynthesis flanked by its native promoter and terminator [41]. In the case of intragenesis, potato is the most recurrently used crop for gene silencing strategies. In fact, the first intragenic application was the increase in amylopectin content in potato by silencing the *granule-bound starch synthase* gene (*GBSS*), responsible for the synthesis of amylose in potato [204]. The silencing construct contains an antisense *GBSS* gene composed of only potato sequences and is controlled by the potato *GBSS* promoter. However, the terminator is the one of the *nopaline synthase* gene (*nos*) from *A. tumefaciens*; thus, this crop could not be considered as completely intragenic. Nevertheless, this potato was released to the field in the EU in 2007 (B/NL/07/04) with the potato *GBSS* terminator, i.e., a fully intragenic potato plant. Another intragenic potato was engineered to reduce the acrylamide content in processed potatoes (without yield penalty or affecting the tuber shape) by silencing one *asparagine synthase* gene (*StAs1*) [205]. The development of other intragenic potatoes was achieved by overexpressing the *lycopene b-ciclase* (*StLYCb*) gene controlled by the potato *GBSS* promoter, which incited β-carotene accumulation in potato tubers [27]. In the case of tomato, carotenoid and flavonoid contents were enhanced simultaneously through the suppression of the *DE-ETIOLATED1* (*DET1*) gene by using RNAi technology and fruit-specific promoters [28]. A gluten-free wheat has also been obtained using this technology by silencing a γ-gliadin gene [206]. The iron content in wheat flour has been increased by more than 2-fold following the expression of a *vacuolar iron transporter* gene (*TaVIT2*) under the control of a wheat endosperm-specific promoter [12]. Finally, Dupont-Pioneer and Monsanto have developed two high oleic soybean oils, Plenish® and Vistive® Gold, respectively, which are currently available in the USA market.

5. Regulation of Plant Breeding Methods

The current regulatory framework could present an obstacle when the above-described techniques are used in crop biofortification, except for conventional breeding, which is not subject to any specific law. However, this is not the case for modern biotechnology techniques. Genetically modified (GM) crops have been demonstrated to be safe countless times, as supported by more than 100 Nobel laureates [231]. In addition, thousands of risk assessments conducted by independent federal regulatory agencies on GM crops have found that there is not different risks between GM and non-GM crops [232]. Nevertheless, there is a widespread lack of acceptance associated with the artificial combination of foreign genetic elements and the use of antibiotic or herbicide resistance selectable markers. All this has triggered alerts about potential health and environmental risks in case gene flow from GM to other non-GM crops [233]. Furthermore, the legislation continues to be strict and differs largely in each country.

In 2019, genetically engineered crops were cultivated in 29 countries, covering a total of 190 million hectares worldwide [232]. North and South America are the biggest producers, followed by Asia, where the law is more flexible. In fact, out of these 190 million hectares of biotech crops cultivation, 174 (90% of the total area) are located in only five countries: USA, Brazil, Argentina, Canada, and India (sorted in descending order) [232]. In the case of the European Union (EU), GMO regulation is one of the most severe, since it assumes that GM crops are intrinsically different (potentially dangerous) [234]. Thus, most countries have used the opt-out clause in relation to the GM crop cultivation and only six countries allow it, having permitted only the cultivation of a GM crop, Bt maize. This led to a decline in research and development (RD) investment in Europa from one-third of the global expenses in agriculture in the mid-1990s to less than 10% by 2013 [235]. Nevertheless, it is worthy to remark that, in England, the rules have been recently relaxed as a consequence of Brexit. Field trials of gene-edited crops with research purposes will be allowed without the current impediments and "red tape", being only necessary to notify it to the Department for Environment, Food, and Rural Affairs (DEFRA) (https://www.gov.uk/government/news, accessed on 3 March 2022). In addition, these measures are likely to be extended to

the rest of UK and a redefinition of the law about genetic modification is also expected. However, until then, gene-edited plants will still be considered GMOs and their commercial cultivation will have to be authorized under the actual law. In many African countries, there is either not any regulatory framework, or it is very restrictive, in spite of being regarded as the part of the world with the largest potential to benefit from the adoption of GM crops due to the high rates of hunger and malnutrition. Notwithstanding, the number of countries embracing GM crops in this continent has been doubled from three in 2018 to six in 2019 [232].

Despite the huge number of developed crops with enhanced traits through genetic engineering, only four different biotech crops cover more than 95% of the cultivated area (soybean, maize, cotton, and canola) and, in most cases, the modified traits are related to herbicide tolerance and insect resistance [232]. Therefore, additional efforts are needed to approve GM crops with enhanced nutritional value in order to contribute to the end of world hunger. Nowadays, transgenesis, cisgenesis, and intragenesis are subject to the same regulation in the vast majority of countries. However, cisgenic and intragenic crops are generally more accepted by the general public and are expected to be regulated less severely in the coming years in some countries [236]. In fact, in Canada, the regulation system is based on the final product rather than on the process to obtain it, which has relaxed the control of these kinds of crops in comparison with the transgenic ones [237]. In Australia, cisgenic plants are not considered GMOs, as stated in Gene Technology Regulations, whereby organisms that are not GMO include "a mutant organism in which the mutational event did not involve the introduction of any foreign nucleic acid" [238]. Other countries are also evaluating cisgenic and intragenic crop regulation. For example, in 2012, the European Food Safety Authority (EFSA) proposed a less precautionary approach to regulate cisgenesis, as it is supposed to entail similar hazards to conventional breeding as introduced by unmodified genes [239]. In the case of intragenesis, the EFSA affirmed that hazards are less predictable due to the recombination of different genetic elements, despite belonging to the same gene pool [240]. However, crops developed by RNAi technology, considered an intragenic approach, have recently received a positive opinion from this organization after determination of risk assessments [240]. In USA, the Environmental Protection Agency (EPA) is also discussing a less strict regulatory framework for cisgenesis and intragenesis approaches, especially when enhanced traits are related to pest resistance [241]. Furthermore, a lot of studies have confirmed a higher consumer and farmer acceptance of cisgenic and intragenic crops than transgenic ones because they are considered to be more natural [227–229]. This, together with the favorable opinions about cisgenesis and intragenesis from public organizations, should pave the way to less stringent regulations for these types of crops. Furthermore, a recent worldwide study has shown that consumers are willing to pay up to 23.9% more for GM-biofortified crops [242].

6. Future Perspectives

The Sustainable Development Goal 2 of the United Nations (UN) consists of ending all forms of hunger, including hidden hunger, before 2030. Nevertheless, projections show that unless serious actions are taken to accelerate the process, hunger will not be eradicated by that year. In fact, current progress is stalled or worsening [1]. Biofortification could substantially help to achieve that objective, as there are cost-effective strategies available. The technologies to explore genomic (i.e., SNP genotyping) and metabolic diversity are evolving astonishingly fast and becoming more and more high-throughput and, at least in the first case, affordable. Similarly, the approaches to identify and introduce the genomic regions responsible for the crop biortification, in this case, are becoming more accurate. However, all of them present some limitations, as we have discussed before. In the case of conventional breeding, the lack of genetic variability and the investment of time, although alleviated by the use of genomic tools (Table 4), to some extent, make it an insufficient strategy to reach the expected food demands [243]. Modern biotechnological techniques

allow us to overcome those hurdles, though they are hampered by regulatory barriers, either non-existing specific laws or especially strict ones, as described in the previous section. Technology is progressing faster than the regulations and this gap is holding us back, for instance, to achieve the UN Sustainable Development Goals.

In parallel, an effort to illuminate the safety of genetically engineered crops in a clear and understandable manner is essential in order to increase their acceptance among the general public and political organizations. It would be also interesting to improve research and development of biofortified crops in developing countries, where malnutrition is a real burden.

7. Conclusions

Considering the expected increase in population in the next years, the challenge is not only to produce enough quantity of food to feed the global population, but also to ensure that food is nutritionally rich to ensure balanced diets. It is well established that biofortification is a cost-effective strategy and a promising approach to fight against global hunger, especially in developing countries. Currently, a large number of biofortified crops have been developed and even released, mainly those obtained through conventional breeding, but also some of them through modern biotechnological techniques. Nevertheless, GMO rejection implies an obstacle and it is frequently based on political preferences in spite of scientific evidences that support the safety of GM-biofortified crops. Here, it is necessary to set aside political and populist views not built on scientific results in order to guarantee food security, a global priority matter. Thus, the likely approval of cisgeneic and intragenic crops, and the less likely but also possible approval of transgenic ones, combined with conventional breeding and genome editing technologies, would place us closer and faster to the zero-hunger goal.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms23063086/s1, Supplementary File S1: Bibliographic search criteria to elaborate the present review.

Author Contributions: Conceptualization, A.D.; writing—original draft preparation, A.D. and I.M.-L.; writing—review and editing, A.D.; supervision and funding acquisition, A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Institute for Agricultural and Food Research and Technology (INIA) through the project RTA2017-00093-00-00, and by the Government of Aragón through the project LMP164_18 and the funds granted to the consolidated research group A12_20R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética". I.M.L. was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU), and the Spanish State Research Agency (AEI).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank J.R. Bertolín from "Laboratorio de Valoración Nutritiva (Department of Animal Science, CITA)" for his expert advice on the analytical methods included in this review.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the writing of the manuscript, or in the decision to publish it.

References

1. FAO; IFAD; UNICEF; WPF; WHO. *The State of Food Security and Nutrition in the World 2021. Trasforming Food System for Food Security, Improved Nutrition and Affordable Healthy Diets for All*, 1st ed.; FAO: Rome, Italy, 2021; ISBN 978-92-5-134325-8.
2. UNICEF. First call for children. In *World Declaration and 1990–2000 Plan of Action on the Survival, Protection and Development of Children*; UNICEF: New York, NY, USA, 1990.

3. Medina-Lozano, I.; Díaz, A. Nutritional Value and Phytochemical Content of Crop Landraces and Traditional Varieties. In *Landraces—Traditional Variety and Natural Breed*; Elkelish, A., Ed.; IntechOpen: Rijeka, Croatia, 2021; pp. 95–116, ISBN 978-1-83968-718-1.
4. Paul, S.; Ali, N.; Datta, S.K.; Datta, K. Development of an Iron-enriched High-yieldings Indica Rice Cultivar by Introgression of A High-iron Trait from Transgenic Iron-biofortified Rice. *Plant Foods Hum. Nutr.* **2014**, *69*, 203–208. [CrossRef] [PubMed]
5. Neelam, K.; Rawat, N.; Tiwari, V.K.; Kumar, S.; Chhuneja, P.; Singh, K.; Randhawa, G.S.; Dhaliwal, H.S. Introgression of group 4 and 7 chromosomes of *Ae. peregrina* in wheat enhances grain iron and zinc density. *Mol. Breed.* **2011**, *28*, 623–634. [CrossRef]
6. Haynes, K.G.; Yencho, G.C.; Clough, M.E.; Henninger, M.R.; Sterrett, S.B. Genetic Variation for Potato Tuber Micronutrient Content and Implications for Biofortification of Potatoes to Reduce Micronutrient Malnutrition. *Am. J. Potato Res.* **2012**, *89*, 192–198. [CrossRef]
7. Blair, M.W.; Astudillo, C.; Grusak, M.A.; Graham, R.; Beebe, S.E. Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* **2009**, *23*, 197–207. [CrossRef]
8. Aluru, M.R.; Rodermel, S.R.; Reddy, M.B. Genetic modification of *low phytic acid 1-1* maize to enhance iron content and bioavailability. *J. Agric. Food Chem.* **2011**, *59*, 12954–12962. [CrossRef] [PubMed]
9. Beasley, J.T.; Bonneau, J.P.; Sánchez-Palacios, J.T.; Moreno-Moyano, L.T.; Callahan, D.L.; Tako, E.; Glahn, R.P.; Lombi, E.; Johnson, A.A.T. Metabolic engineering of bread wheat improves grain iron concentration and bioavailability. *Plant Biotechnol. J.* **2019**, *17*, 1514–1526. [CrossRef] [PubMed]
10. Narayanan, N.; Beyene, G.; Chauhan, R.D.; Gaitán-Solís, E.; Gehan, J.; Butts, P.; Siritunga, D.; Okwuonu, I.; Woll, A.; Jiménez-Aguilar, D.M.; et al. Biofortification of field-grown cassava by engineering expression of an iron transporter and ferritin. *Nat. Biotechnol.* **2019**, *37*, 144–151. [CrossRef]
11. Nozoye, T.; Otani, M.; Senoura, T.; Nakanishi, H.; Nishizawa, N.K. Overexpression of barley *nicotianamine synthase 1* confers tolerance in the sweet potato to iron deficiency in calcareous soil. *Plant Soil* **2017**, *418*, 75–88. [CrossRef]
12. Connorton, J.M.; Jones, E.R.; Rodríguez-Ramiro, I.; Fairweather-Tait, S.; Uauy, C.; Balk, J. Wheat Vacuolar Iron Transporter TaVIT2 Transports Fe and Mn and Is Effective for Biofortification. *Plant Physiol.* **2017**, *174*, 2434–2444. [CrossRef]
13. Menkir, A.; Maziya-Dixon, B.; Mengesha, W.; Rocheford, T.; Alamu, E.O. Accruing genetic gain in pro-vitamin A enrichment from harnessing diverse maize germplasm. *Euphytica* **2017**, *213*, 105. [CrossRef]
14. Hossain, F.; Muthusamy, V.; Zunjare, R.U. Molecular Breeding for Development of Biofortified Maize Hybrids in India. In Proceedings of the Extended Summaries: 13th Asian Maize Conference on and Expert Consultation on Maize for Food, Feed, Nutrition and Environmental Security, Ludhiana, India, 8–10 October 2018; pp. 220–230.
15. Ceballos, H.; Morante, N.; Sanchez, T.; Ortiz, D.; Aragón, I.; Chávez, A.; Pizarro, M.; Calle, F.; Dominique, D. Rapid Cycling Recurrent Selection for Increased Carotenoids Content in Cassava Roots. *Crop Sci.* **2013**, *53*, 2342–2351. [CrossRef]
16. Rezaei, M.; Deokar, A.; Arganosa, G.; Roorkiwal, M.; Pandey, S.; Warkentin, T.; Varshney, R.; Tar'an, B. Mapping Quantitative Trait Loci for Carotenoid Concentration in Three F Populations of Chickpea. *Plant Genome* **2019**, *12*, 1–12. [CrossRef] [PubMed]
17. Ye, X.; Al-Babili, S.; Klöti, A.; Zhang, J.; Lucca, P.; Beyer, P.; Potrykus, I. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **2000**, *287*, 303–305. [CrossRef] [PubMed]
18. Broad, R.C.; Bonneau, J.P.; Beasley, J.T.; Roden, S.; Sadowski, P.; Jewell, N.; Brien, C.; Berger, B.; Tako, E.; Glahn, R.P.; et al. Effect of Rice GDP-L-Galactose Phosphorylase Constitutive Overexpression on Ascorbate Concentration, Stress Tolerance, and Iron Bioavailability in Rice. *Front. Plant Sci.* **2020**, *11*, 595439. [CrossRef] [PubMed]
19. Aluru, M.; Xu, Y.; Guo, R.; Wang, Z.; Li, S.; White, W.; Wang, K.; Rodermel, S. Generation of transgenic maize with enhanced provitamin A content. *J. Exp. Bot.* **2008**, *59*, 3551–3562. [CrossRef] [PubMed]
20. Liang, Q.; Wang, K.; Liu, X.; Riaz, B.; Jiang, L.; Wan, X.; Ye, X.; Zhang, C. Improved folate accumulation in genetically modified maize and wheat. *J. Exp. Bot.* **2019**, *70*, 1539–1551. [CrossRef] [PubMed]
21. De Lepeleire, J.; Strobbe, S.; Verstraete, J.; Blancquaert, D.; Ambach, L.; Visser, R.G.F.; Stove, C.; Van Der Straeten, D. Folate Biofortification of Potato by Tuber-Specific Expression of Four Folate Biosynthesis Genes. *Mol. Plant* **2018**, *11*, 175–188. [CrossRef] [PubMed]
22. Yazdani, M.; Sun, Z.; Yuan, H.; Zeng, S.; Thannhauser, T.W.; Vrebalov, J.; Ma, Q.; Xu, Y.; Fei, Z.; Van Eck, J.; et al. Ectopic expression of ORANGE promotes carotenoid accumulation and fruit development in tomato. *Plant Biotechnol. J.* **2019**, *17*, 33–49. [CrossRef] [PubMed]
23. Zhang, Y.; Butelli, E.; Alseekh, S.; Tohge, T.; Rallapalli, G.; Luo, J.; Kawar, P.G.; Hill, L.; Santino, A.; Fernie, A.R.; et al. Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* **2015**, *6*, 8635. [CrossRef]
24. Paul, J.-Y.; Khanna, H.; Kleidon, J.; Hoang, P.; Geijskes, J.; Daniells, J.; Zaplin, E.; Rosenberg, Y.; James, A.; Mlalazi, B.; et al. Golden bananas in the field: Elevated fruit pro-vitamin A from the expression of a single banana transgene. *Plant Biotechnol. J.* **2017**, *15*, 520–532. [CrossRef]
25. Ramírez Rivera, N.G.; García-Salinas, C.; Aragão, F.J.L.; Díaz de la Garza, R.I. Metabolic engineering of folate and its precursors in Mexican common bean (*Phaseolus vulgaris* L.). *Plant Biotechnol. J.* **2016**, *14*, 2021–2032. [CrossRef] [PubMed]
26. Yu, B.; Lydiate, D.J.; Young, L.W.; Schäfer, U.A.; Hannoufa, A. Enhancing the carotenoid content of *Brassica napus* seeds by downregulating lycopene epsilon cyclase. *Transgenic Res.* **2008**, *17*, 573–585. [CrossRef] [PubMed]
27. Song, X.; Zhu, W.; Tang, R.; Cai, J.; Chen, M.; Yang, Q. Over-expression of *StLCYb* increases β-carotene accumulation in potato tubers. *Plant Biotechnol. Rep.* **2016**, *10*, 95–104. [CrossRef]

28. Davuluri, G.R.; van Tuinen, A.; Fraser, P.D.; Manfredonia, A.; Newman, R.; Burgess, D.; Brummell, D.A.; King, S.R.; Palys, J.; Uhlig, J.; et al. Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. *Nat. Biotechnol.* **2005**, *23*, 890–895. [CrossRef] [PubMed]
29. Singh, S.P.; Gruissem, W.; Bhullar, N.K. Single genetic locus improvement of iron, zinc and β-carotene content in rice grains. *Sci. Rep.* **2017**, *7*, 6883. [CrossRef] [PubMed]
30. Ruiz-López, N.; Haslam, R.P.; Venegas-Calerón, M.; Larson, T.R.; Graham, I.A.; Napier, J.A.; Sayanova, O. The synthesis and accumulation of stearidonic acid in transgenic plants: A novel source of "heart-healthy" omega-3 fatty acids. *Plant Biotechnol. J.* **2009**, *7*, 704–716. [CrossRef] [PubMed]
31. Cheng, B.; Wu, G.; Vrinten, P.; Falk, K.; Bauer, J.; Qiu, X. Towards the production of high levels of eicosapentaenoic acid in transgenic plants: The effects of different host species, genes and promoters. *Transgenic Res.* **2010**, *19*, 221–229. [CrossRef]
32. Jing, G.; Tang, D.; Yao, Y.; Su, Y.; Shen, Y.; Bai, Y.; Jing, W.; Zhang, Q.; Lin, F.; Guo, D.; et al. Seed specifically over-expressing *DGAT2A* enhances oil and linoleic acid contents in soybean seeds. *Biochem. Biophys. Res. Commun.* **2021**, *568*, 143–150. [CrossRef]
33. Liu, Q.; Guo, Q.; Akbar, S.; Zhi, Y.; El Tahchy, A.; Mitchell, M.; Li, Z.; Shrestha, P.; Vanhercke, T.; Ral, J.P.; et al. Genetic enhancement of oil content in potato tuber (*Solanum tuberosum* L.) through an integrated metabolic engineering strategy. *Plant Biotechnol. J.* **2017**, *15*, 56–67. [CrossRef]
34. Kim, W.-S.; Sun-Hyung, J.; Oehrle, N.W.; Jez, J.M.; Krishnan, H.B. Overexpression of ATP sulfurylase improves the sulfur amino acid content, enhances the accumulation of Bowman-Birk protease inhibitor and suppresses the accumulation of the β-subunit of β-conglycinin in soybean seeds. *Sci. Rep.* **2020**, *10*, 14989. [CrossRef] [PubMed]
35. Tang, M.; He, X.; Luo, Y.; Ma, L.; Tang, X.; Huang, K. Nutritional assessment of transgenic lysine-rich maize compared with conventional quality protein maize. *J. Sci. Food Agric.* **2013**, *93*, 1049–1054. [CrossRef] [PubMed]
36. Yang, Q.Q.; Zhang, C.Q.; Chan, M.L.; Zhao, D.S.; Chen, J.Z.; Wang, Q.; Li, Q.F.; Yu, H.X.; Gu, M.H.; Sun, S.S.M.; et al. Biofortification of rice with the essential amino acid lysine: Molecular characterization, nutritional evaluation, and field performance. *J. Exp. Bot.* **2016**, *67*, 4285–4296. [CrossRef] [PubMed]
37. Xu, M.; Zhao, S.; Zhang, Y.; Yin, H.; Peng, X.; Cheng, Z.; Yang, Z.; Zheng, J. Production of Marker-Free Transgenic Rice (*Oryza sativa* L.) with Improved Nutritive Quality Expressing *AmA1*. *Iran. J. Biotechnol.* **2017**, *15*, 102–110. [CrossRef] [PubMed]
38. Garg, M.; Chawla, M.; Chunduri, V.; Kumar, R.; Sharma, S.; Sharma, N.K.; Kaur, N.; Kumar, A.; Mundey, J.K.; Saini, M.K.; et al. Transfer of grain colors to elite wheat cultivars and their characterization. *J. Cereal Sci.* **2016**, *71*, 138–144. [CrossRef]
39. Ingallina, C.; Maccelli, A.; Spano, M.; Di Matteo, G.; Di Sotto, A.; Giusti, A.M.; Vinci, G.; Di Giacomo, S.; Rapa, M.; Ciano, S.; et al. Chemico-biological characterization of torpedino di fondi® tomato fruits: A comparison with san marzano cultivar at two ripeness stages. *Antioxidants* **2020**, *9*, 1027. [CrossRef] [PubMed]
40. Zhu, Q.; Zeng, D.; Yu, S.; Cui, C.; Li, J.; Li, H.; Chen, J.; Zhang, R.; Zhao, X.; Chen, L.; et al. From Golden Rice to aSTARice: Bioengineering Astaxanthin Biosynthesis in Rice Endosperm. *Mol. Plant* **2018**, *11*, 1440–1448. [CrossRef] [PubMed]
41. Krens, F.A.; Schaart, J.G.; van der Burgh, A.M.; TinnenbroekCapel, I.E.M.; Groenwold, R.; Kodde, L.P.; Broggini, G.A.L.; Gessler, C.; Schouten, H.J. Cisgenic apple trees; development, characterization, and performance. *Front. Plant Sci.* **2015**, *6*, 286. [CrossRef]
42. Hoddinott, J.F.; Rosegrant, M.W.; Torero, M. Investments to reduce hunger and undernutrition. In *Global Problems, Smart Solutions*; Lomborg, B., Ed.; Cambridge University Press: Cambridge, UK, 2013; pp. 332–367.
43. Graham, R.D.; Welch, R.M. *Breeding for Staple-Food Crops with High Micronutrient Density: Working Papers on Agricultural Strategies for Micronutrients*, 3rd ed.; International Food Policy Institute: Washington, DC, USA, 1996.
44. Graham, R.D.; Welch, R.M.; Bouis, H.E. Addressing Micronutrient Malnutrition Through Enhancing the Nutritional Quality of Staple Foods: Principles, Perspectives and Knowledge Gaps. *Adv. Agron.* **2001**, *70*, 77–142. [CrossRef]
45. United Nations; Department of Economic and Social; Affairs, P.D. *World Population Prospects, Medium Prognosis*; The 2019 Revision: New York, NY, USA, 2019.
46. Wortmann, L.; Enneking, U.; Daum, D. German Consumers' Attitude Towards Selenium-Biofortified Apples and Acceptance of Related Nutrition and Health Claims. *Nutrients* **2018**, *10*, 190. [CrossRef]
47. Welk, A.K.; Kleine-kalmer, R.; Daum, D.; Enneking, U. Consumer Acceptance and Market Potential of Iodine-Biofortified Fruit and Vegetables in Germany. *Nutrients* **2021**, *13*, 4198. [CrossRef] [PubMed]
48. Adekambi, S.A.; Okello, J.J.; Rajendran, S.; Acheremu, K.; Carey, E.E.; Low, J.; Abidin, P.E. Effect of varietal attributes on the adoption of an orange-fleshed sweetpotato variety in Upper East and Northern Ghana. *Outlook Agric.* **2020**, *49*, 311–320. [CrossRef] [PubMed]
49. Gannon, B.; Kaliwile, C.; Arscott, S.A.; Schmaelzle, S.; Chileshe, J.; Kalungwana, N.; Mosonda, M.; Pixley, K.; Masi, C.; Tanumihardjo, S.A. Biofortified orange maize is as efficacious as a vitamin A supplement in Zambian children even in the presence of high liver reserves of vitamin A: A community-based, randomized placebo-controlled trial. *Am. J. Clin. Nutr.* **2014**, *100*, 1541–1550. [CrossRef] [PubMed]
50. Komar, A.A. Single Nucleotide Polymorphisms. *Methods Mol. Biol.* **2009**, *578*, 23–39.
51. Collard, B.C.Y.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **2005**, *142*, 169–196. [CrossRef]
52. Unterseer, S.; Bauer, E.; Haberer, G.; Seidel, M.; Knaak, C.; Ouzunova, M.; Meitinger, T.; Strom, T.M.; Fries, R.; Pausch, H.; et al. A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* **2014**, *15*, 1–15. [CrossRef] [PubMed]

53. Xu, C.; Ren, Y.; Jian, Y.; Guo, Z.; Zhang, Y.; Xie, C.; Fu, J.; Wang, H.; Wang, G.; Xu, Y.; et al. Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* **2017**, *37*, 1–12. [CrossRef] [PubMed]
54. Tung, C.W.; Zhao, K.; Wright, M.H.; Ali, M.L.; Jung, J.; Kimball, J.; Tyagi, W.; Thomson, M.J.; McNally, K.; Leung, H.; et al. Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.). *Rice* **2010**, *3*, 205–217. [CrossRef]
55. Singh, N.; Jayaswal, P.K.; Panda, K.; Mandal, P.; Kumar, V.; Singh, B.; Mishra, S.; Singh, Y.; Singh, R.; Rai, V.; et al. Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* **2015**, *5*, 1–9. [CrossRef]
56. Bauer, E.; Schmutzer, T.; Barilar, I.; Mascher, M.; Gundlach, H.; Martis, M.M.; Twardziok, S.O.; Hackauf, B.; Gordillo, A.; Wilde, P.; et al. Towards a whole-genome sequence for rye (*Secale cereale* L.). *Plant J.* **2017**, *89*, 853–869. [CrossRef]
57. Winfield, M.O.; Allen, A.M.; Burridge, A.J.; Barker, G.L.A.; Benbow, H.R.; Wilkinson, P.A.; Coghill, J.; Waterfall, C.; Davassi, A.; Scopes, G.; et al. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* **2016**, *14*, 1195–1206. [CrossRef] [PubMed]
58. Allen, A.M.; Winfield, M.O.; Burridge, A.J.; Downie, R.C.; Benbow, H.R.; Barker, G.L.A.; Wilkinson, P.A.; Coghill, J.; Waterfall, C.; Davassi, A.; et al. Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol. J.* **2017**, *15*, 390–401. [CrossRef] [PubMed]
59. Karadi, A.; Samineni, S.; Sajja, S.; Sharma, M.; Thudi, M.; Mallikarjuna, B.P.; Viswanatha, K.P.; Varshney, R.K.; Gaur, P.M. Molecular mapping of dry root rot resistance genes in chickpea (*Cicer arietinum* L.). *Euphytica* **2021**, *217*, 1–13. [CrossRef]
60. Stoffel, K.; van Leeuwen, H.; Kozik, A.; Caldwell, D.; Ashrafi, H.; Cui, X.; Tan, X.; Hill, T.; Reyes-Chin-Wo, S.; Truco, M.J.; et al. Development and application of a 6.5 million feature Affymetrix Genechip® for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics* **2012**, *13*, 1–17. [CrossRef] [PubMed]
61. Vos, P.G.; Uitdewilligen, J.G.A.M.L.; Voorrips, R.E.; Visser, R.G.F.; van Eck, H.J. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): An insight into the breeding history. *Theor. Appl. Genet.* **2015**, *128*, 2387–2401. [CrossRef] [PubMed]
62. Lee, Y.G.; Jeong, N.; Kim, J.H.; Lee, K.; Kim, K.H.; Pirani, A.; Ha, B.K.; Kang, S.T.; Park, B.S.; Moon, J.K.; et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* **2015**, *81*, 625–636. [CrossRef]
63. Bassil, N.V.; Davis, T.M.; Zhang, H.; Ficklin, S.; Mittmann, M.; Webster, T.; Mahoney, L.; Wood, D.; Alperin, E.S.; Rosyara, U.R.; et al. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry Fragaria × ananassa. *BMC Genomics* **2015**, *16*, 1–30. [CrossRef]
64. Bianco, L.; Cestaro, A.; Linsmith, G.; Muranty, H.; Denancé, C.; Théron, A.; Poncet, C.; Micheletti, D.; Kerschbamer, E.; Di Pierro, E.A.; et al. Development and validation of the Axiom®Apple480K SNP genotyping array. *Plant J.* **2016**, *86*, 62–74. [CrossRef]
65. Pandey, M.K.; Agarwal, G.; Kale, S.M.; Clevenger, J.; Nayak, S.N.; Sriswathi, M.; Chitikineni, A.; Chavarro, C.; Chen, X.; Upadhyaya, H.D.; et al. Development and Evaluation of a High Density Genotyping "Axiom-*Arachis*" Array with 58 K SNPs for Accelerating Genetics and Breeding in Groundnut. *Sci. Rep.* **2017**, *7*, 1–10. [CrossRef] [PubMed]
66. Comadran, J.; Kilian, B.; Russell, J.; Ramsay, L.; Stein, N.; Ganal, M.; Shaw, P.; Bayer, M.; Thomas, W.; Marshall, D.; et al. Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **2012**, *44*, 1388–1391. [CrossRef]
67. Ganal, M.W.; Durstewitz, G.; Polley, A.; Bérard, A.; Buckler, E.S.; Charcosset, A.; Clarke, J.D.; Graner, E.M.; Hansen, M.; Joets, J.; et al. A Large Maize (*Zea mays* L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. *PLoS ONE* **2011**, *6*, e28334. [CrossRef] [PubMed]
68. Rousselle, Y.; Jones, E.; Charcosset, A.; Moreau, P.; Robbins, K.; Stich, B.; Knaak, C.; Flament, P.; Karaman, Z.; Martinant, J.P.; et al. Study on Essential Derivation in Maize: III. Selection and Evaluation of a Panel of Single Nucleotide Polymorphism Loci for Use in European and North American Germplasm. *Crop Sci.* **2015**, *55*, 1170–1180. [CrossRef]
69. Tinker, N.A.; Chao, S.; Lazo, G.R.; Oliver, R.E.; Huang, Y.; Poland, J.A.; Jellen, E.N.; Maughan, P.J.; Kilian, A.; Jackson, E.W. A SNP Genotyping Array for Hexaploid Oat. *Plant Genome* **2014**, *7*, 1–8. [CrossRef]
70. Chen, W.; Gao, Y.; Xie, W.; Gong, L.; Lu, K.; Wang, W.; Li, Y.; Liu, X.; Zhang, H.; Dong, H.; et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **2014**, *46*, 714–721. [CrossRef] [PubMed]
71. Yu, H.; Xie, W.; Li, J.; Zhou, F.; Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* **2014**, *12*, 28–37. [CrossRef] [PubMed]
72. Cavanagh, C.R.; Chao, S.; Wang, S.; Huang, B.E.; Stephen, S.; Kiani, S.; Forrest, K.; Saintenac, C.; Brown-Guedira, G.L.; Akhunova, A.; et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8057–8062. [CrossRef] [PubMed]
73. Clarke, W.E.; Higgins, E.E.; Plieske, J.; Wieseke, R.; Sidebottom, C.; Khedikar, Y.; Batley, J.; Edwards, D.; Meng, J.; Li, R.; et al. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* **2016**, *129*, 1887–1899. [CrossRef] [PubMed]
74. Bachlava, E.; Taylor, C.A.; Tang, S.; Bowers, J.E.; Mandel, J.R.; Burke, J.M.; Knapp, S.J. SNP Discovery and Development of a High-Density Genotyping Array for Sunflower. *PLoS ONE* **2012**, *7*, e29814. [CrossRef] [PubMed]

75. Livaja, M.; Unterseer, S.; Erath, W.; Lehermeier, C.; Wieseke, R.; Plieske, J.; Polley, A.; Luerßen, H.; Wieckhorst, S.; Mascher, M.; et al. Diversity analysis and genomic prediction of *Sclerotinia* resistance in sunflower using a new 25 K SNP genotyping array. *Theor. Appl. Genet.* **2016**, *129*, 317–329. [CrossRef] [PubMed]
76. Close, T.J.; Lucas, M.R.; Muñoz-Amatriain, M.; Mirebrahim, H.; Wanamaker, S.; Barkley, N.A.; Clair, S.S.; Guo, Y.-N.; Lo, S.; Huynh, B.L. A new SNP-genotyping resource for cowpea and its deployment for breeding. In Proceedings of the Plant and Animal Genome Conference, San Diego, CA, USA, 10–14 January 2015; Volume 23, p. P0784.
77. Hamilton, J.P.; Hansey, C.N.; Whitty, B.R.; Stoffel, K.; Massa, A.N.; Van Deynze, A.; De Jong, W.S.; Douches, D.S.; Buell, C.R. Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics* **2011**, *12*, 1–11. [CrossRef]
78. Sim, S.C.; Durstewitz, G.; Plieske, J.; Wieseke, R.; Ganal, M.W.; van Deynze, A.; Hamilton, J.P.; Buell, C.R.; Causse, M.; Wijeratne, S.; et al. Development of a Large SNP Genotyping Array and Generation of High-Density Genetic Maps in Tomato. *PLoS ONE* **2012**, *7*, e40563. [CrossRef] [PubMed]
79. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE* **2013**, *8*, e54985. [CrossRef] [PubMed]
80. Bianco, L.; Cestaro, A.; Sargent, D.J.; Banchi, E.; Derdak, S.; Di Guardo, M.; Salvi, S.; Jansen, J.; Viola, R.; Gut, I.; et al. Development and validation of a 20K Single Nucleotide Polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica Borkh*). *PLoS ONE* **2014**, *9*, e110377. [CrossRef] [PubMed]
81. Chagné, D.; Crowhurst, R.N.; Troggio, M.; Davey, M.W.; Gilmore, B.; Lawley, C.; Vanderzande, S.; Hellens, R.P.; Kumar, S.; Cestaro, A.; et al. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. *PLoS ONE* **2012**, *7*, e31745. [CrossRef] [PubMed]
82. Peace, C.; Bassil, N.; Main, D.; Ficklin, S.; Rosyara, U.R.; Stegmeir, T.; Sebolt, A.; Gilmore, B.; Lawley, C.; Mockler, T.C.; et al. Development and Evaluation of a Genome-Wide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry. *PLoS ONE* **2012**, *7*, e48305. [CrossRef] [PubMed]
83. Verde, I.; Bassil, N.; Scalabrin, S.; Gilmore, B.; Lawley, C.T.; Gasic, K.; Micheletti, D.; Rosyara, U.R.; Cattonaro, F.; Vendramin, E.; et al. Development and Evaluation of a 9k SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm. *PLoS ONE* **2012**, *7*, e35668. [CrossRef]
84. Micheletti, D.; Dettori, M.T.; Micali, S.; Aramini, V.; Pacheco, I.; Da Silva Linge, C.; Foschi, S.; Banchi, E.; Barreneche, T.; Quilot-Turion, B.; et al. Whole-Genome Analysis of Diversity and SNP-Major Gene Association in Peach Germplasm. *PLoS ONE* **2015**, *10*, e0136803. [CrossRef]
85. Mas-Gómez, J.; Cantín, C.M.; Moreno, M.; Prudencio, Á.S.; Gómez-Abajo, M.; Bianco, L.; Troggio, M.; Martínez-Gómez, P.; Rubio, M.; Martínez-García, P.J. Exploring genome-wide diversity in the national peach (*Prunus persica*) germplasm collection at CITA (Zaragoza, Spain). *Agronomy* **2021**, *11*, 481. [CrossRef]
86. Montanari, S.; Saeed, M.; Knäbel, M.; Kim, Y.K.; Troggio, M.; Malnoy, M.; Velasco, R.; Fontana, P.; Won, K.H.; Durel, C.E.; et al. Identification of Pyrus Single Nucleotide Polymorphisms (SNPs) and Evaluation for Genetic Mapping in European Pear and Interspecific Pyrus Hybrids. *PLoS ONE* **2013**, *8*, e77022. [CrossRef]
87. Myles, S.; Chia, J.M.; Hurwitz, B.; Simon, C.; Zhong, G.Y.; Buckler, E.; Ware, D. Rapid genomic characterization of the genus *Vitis*. *PLoS ONE* **2010**, *5*, e8219. [CrossRef]
88. Le Paslier, M.-C.; Choisne, N.; Scalabrin, S.; Bacilieri, R.; Berard, A.; Bounon, R.; Boursiquot, J.-M.; Bras, M.; Brunel, D.; Chauveau, A.; et al. The GrapeReSeq 18K Vitis genotyping chip. IX International Symposium on Grapevine Physiology Biotechnology, La Serena, Chile, 2–26 April 2013; p. 18.
89. De Jong, W.S.; De Jong, D.M.; Bodis, M. A fluorogenic 5′ nuclease (TaqMan) assay to assess dosage of a marker tightly linked to red skin color in autotetraploid potato. *Theor. Appl. Genet.* **2003**, *107*, 1384–1390. [CrossRef]
90. Wu, J.; Wang, Q.; Kang, Z.; Liu, S.; Li, H.; Mu, J.; Dai, M.; Han, D.; Zeng, Q.; Chen, X. Development and validation of KASP-SNP markers for QTL underlying resistance to stripe rust in common wheat cultivar P10057. *Plant Dis.* **2017**, *101*, 2079–2087. [CrossRef] [PubMed]
91. Qureshi, N.; Kandiah, P.; Gessese, M.K.; Nsabiyera, V.; Wells, V.; Babu, P.; Wong, D.; Hayden, M.; Bariana, H.; Bansal, U. Development of co-dominant KASP markers co-segregating with Ug99 effective stem rust resistance gene *Sr26* in wheat. *Mol. Breed.* **2018**, *38*, 1–9. [CrossRef]
92. Collins, D.; Emebiri, L.; Tan, M.K.; El Bouhssini, M.; Wildman, O. Association of KASP markers with Hessian fly resistance in wheat of diverse origin. *Euphytica* **2018**, *214*, 1–8. [CrossRef]
93. Wang, R.; Liu, Y.; Isham, K.; Zhao, W.; Wheeler, J.; Klassen, N.; Hu, Y.; Bonman, J.M.; Chen, J. QTL identification and KASP marker development for productive tiller and fertile spikelet numbers in two high-yielding hard white spring wheat cultivars. *Mol. Breed.* **2018**, *38*, 1–12. [CrossRef] [PubMed]
94. Singh, L.; Anderson, J.A.; Chen, J.; Gill, B.S.; Tiwari, V.K.; Rawat, N. Development and validation of a perfect KASP marker for fusarium head blight resistance gene *Fhb1* in wheat. *Plant Pathol. J.* **2019**, *35*, 200–207. [CrossRef] [PubMed]
95. Fang, T.; Lei, L.; Li, G.; Powers, C.; Hunger, R.M.; Carver, B.F.; Yan, L. Development and deployment of KASP markers for multiple alleles of *Lr34* in wheat. *Theor. Appl. Genet.* **2020**, *133*, 2183–2195. [CrossRef] [PubMed]
96. Grewal, S.; Othmeni, M.; Walker, J.; Hubbart-Edwards, S.; Yang, C.Y.; Scholefield, D.; Ashling, S.; Isaac, P.; King, I.P.; King, J. Development of Wheat-*Aegilops caudata* Introgression Lines and Their Characterization Using Genome-Specific KASP Markers. *Front. Plant Sci.* **2020**, *11*, 606. [CrossRef] [PubMed]

97. Makhoul, M.; Rambla, C.; Voss-Fels, K.P.; Hickey, L.T.; Snowdon, R.J.; Obermeier, C. Overcoming polyploidy pitfalls: A user guide for effective SNP conversion into KASP markers in wheat. *Theor. Appl. Genet.* **2020**, *133*, 2413–2430. [CrossRef] [PubMed]
98. Zhang, S.; Fan, C.; Luo, J.; Huang, L.; Xie, D.; Li, Y.; Chen, Z.; Jiang, B.; Ning, S.; Yuan, Z.; et al. KASP markers to detect sub-chromosomal arm translocations between 6VS of *Haynaldia villosa* and 6AS of wheat. *Euphytica* **2021**, *217*, 10. [CrossRef]
99. Xu, X.; Li, G.; Bai, G.; Bernardo, A.; Carver, B.F.; Amand, P.S.; Armstrong, J.S. Development of KASP markers for wheat greenbug resistance gene *Gb5*. *Crop Sci.* **2021**, *61*, 490–499. [CrossRef]
100. Sangha, J.; Tucker, J.R.; Legge, W.G.; Badea, A. Use of KASP assays for the analysis of rpg4/Rpg5 gene complex for marker-assisted selection for Ug99 stem rust resistance in barley. *Can. J. Plant Pathol.* **2017**, *39*, 578.
101. Steele, K.A.; Quinton-Tulloch, M.J.; Amgai, R.B.; Dhakal, R.; Khatiwada, S.P.; Vyas, D.; Heine, M.; Witcombe, J.R. Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of *indica* rice. *Mol. Breed.* **2018**, *38*, 1–13. [CrossRef] [PubMed]
102. Yang, G.; Chen, S.; Chen, L.; Sun, K.; Huang, C.; Zhou, D.; Huang, Y.; Wang, J.; Liu, Y.; Wang, H.; et al. Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* **2019**, *12*, 1–18. [CrossRef] [PubMed]
103. Yang, Y.; Basnet, B.R.; Ibrahim, A.M.H.; Rudd, J.C.; Chen, X.; Bowden, R.L.; Xue, Q.; Wang, S.; Johnson, C.D.; Metz, R.; et al. Developing KASP Markers on a Major Stripe Rust Resistance QTL in a Popular Wheat TAM 111 Using 90K Array and Genotyping-by-Sequencing SNPs. *Crop Sci.* **2019**, *59*, 165–175. [CrossRef]
104. Addison, C.K.; Angira, B.; Kongchum, M.; Harrell, D.L.; Baisakh, N.; Linscombe, S.D.; Famoso, A.N. Characterization of Haplotype Diversity in the *BADH2* Aroma Gene and Development of a KASP SNP Assay for Predicting Aroma in U.S. Rice. *Rice* **2020**, *13*, 1–9. [CrossRef] [PubMed]
105. Burow, G.; Chopra, R.; Hughes, H.; Xin, Z.; Burke, J. Marker Assisted Selection in Sorghum Using KASP Assay for the Detection of Single Nucleotide Polymorphism/Insertion Deletion. *Methods Mol. Biol.* **2019**, *1931*, 75–84. [CrossRef] [PubMed]
106. Grimm, K.D.S.; Porter, L.D. Development and Validation of KASP Markers for the Identification of *Pea seedborne* mosaic virus Pathotype P1 Resistance in *Pisum sativum*. *Plant Dis.* **2020**, *104*, 1824–1830. [CrossRef]
107. Legendre, R.; McGregor, C. KASP (TM) Markers for Selection for Fruit Shape in Watermelon. *Hortscience* **2019**, *54*, S399.
108. Paudel, L.; Clevenger, J.; McGregor, C. Refining of the egusi locus in watermelon using KASP assays. *Sci. Hortic.* **2019**, *257*, 108665. [CrossRef]
109. Zanotto, S.; Vandenberg, A.; Khazaei, H. Development and validation of a robust KASP marker for *zt2* locus in faba bean (*Vicia faba*). *Plant Breed.* **2020**, *139*, 375–380. [CrossRef]
110. Devran, Z.; Göknur, A.; Mesci, L. Development of molecular markers for the *Mi-1* gene in tomato using the KASP genotyping assay. *Hortic. Environ. Biotechnol.* **2016**, *57*, 156–160. [CrossRef]
111. Devran, Z.; Kahveci, E. Development and validation of a user-friendly KASP marker for the *Sw-5* locus in tomato. *Australas. Plant Pathol.* **2019**, *48*, 503–507. [CrossRef]
112. Han, F.; Zhang, X.; Yuan, K.; Fang, Z.; Yang, L.; Zhuang, M.; Zhang, Y.; Wang, Y.; Liu, Y.; Li, Z.; et al. A user-friendly KASP molecular marker developed for the DGMS-based breeding system in *Brassica oleracea* species. *Mol. Breed.* **2019**, *39*, 1–7. [CrossRef]
113. da Cruz, M.F.A.; Bueno, R.D.; de Souza, F.B.; Moreira, M.A.; de Barros, E.G. Identification of SNPs for fatty acid content in soybean by the HRM technique. *Pesqui. Agropecu. Bras.* **2013**, *48*, 1596–1600. [CrossRef]
114. Rai, V.P.; Singh, A.K.; Jaiswal, H.K.; Singh, S.P.; Singh, R.P.; Waza, S.A. Evaluation of molecular markers linked to fragrance and genetic diversity in Indian aromatic rice. *Turk. J. Botany* **2015**, *39*, 209–217. [CrossRef]
115. Noh, Y.H.; Lee, S.; Whitaker, V.M.; Cearley, K.R.; Cha, J.S. A high-throughput marker-assisted selection system combining rapid DNA extraction and high-resolution melting and simple sequence repeat analysis: Strawberry as a model for fruit crops. *J. Berry Res.* **2017**, *7*, 23–31. [CrossRef]
116. Geng, L.; Li, M.; Xie, S.; Wu, D.; Ye, L.; Zhang, G. Identification of genetic loci and candidate genes related to β-glucan content in barley grain by genome-wide association study in International Barley Core Selected Collection. *Mol. Breed.* **2021**, *41*, 1–12. [CrossRef]
117. Paris, M.; Jones, M.G.K.; Eglinton, J.K. Genotyping Single Nucleotide Polymorphisms for Selection of Barley β-amylase Alleles. *Plant Mol. Biol. Report.* **2002**, *20*, 149–159. [CrossRef]
118. Masouleh, A.K.; Waters, D.L.E.; Reinke, R.F.; Henry, R.J. A high-throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF mass spectrometry. *Plant Biotechnol. J.* **2009**, *7*, 355–363. [CrossRef]
119. Cheng, P.; Holdsworth, W.; Ma, Y.; Coyne, C.J.; Mazourek, M.; Grusak, M.A.; Fuchs, S.; McGee, R.J. Association mapping of agronomic and quality traits in USDA pea single-plant collection. *Mol. Breed.* **2015**, *35*, 1–13. [CrossRef]
120. Lum, H.-K.; Lee, C.-H.; Butt, Y.K.-C.; Lo, S.C.-L. Sodium nitroprusside affects the level of photosynthetic enzymes and glucose metabolism in *Phaseolus aureus* (mung bean). *Nitric Oxide* **2005**, *12*, 220–230. [CrossRef] [PubMed]
121. Salazar, J.A.; Pacheco, I.; Shinya, P.; Zapata, P.; Silva, C.; Aradhya, M.; Velasco, D.; Ruiz, D.; Martínez-Gómez, P.; Infante, R. Genotyping by Sequencing for SNP-Based Linkage Analysis and Identification of QTLs Linked to Fruit Quality Traits in Japanese Plum (*Prunus salicina* Lindl.). *Front. Plant Sci.* **2017**, *8*, 1–14. [CrossRef] [PubMed]
122. Larsen, B.; Migicovsky, Z.; Jeppesen, A.A.; Gardner, K.M.; Toldam-Andersen, T.B.; Myles, S.; Ørgaard, M.; Petersen, M.A.; Pedersen, C. Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition, and Harvest Date. *Plant Genome* **2019**, *12*, 180104. [CrossRef] [PubMed]

123. Pereira, L.; Ruggieri, V.; Pérez, S.; Alexiou, K.G.; Fernández, M.; Jahrmann, T.; Pujol, M.; Garcia-Mas, J. QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biol.* **2018**, *18*, 1–17. [CrossRef] [PubMed]
124. Guo, R.; Dhliwayo, T.; Mageto, E.K.; Palacios-Rojas, N.; Lee, M.; Yu, D.; Ruan, Y.; Zhang, A.; San Vicente, F.; Olsen, M.; et al. Genomic Prediction of Kernel Zinc Concentration in Multiple Maize Populations Using Genotyping-by-Sequencing and Repeat Amplification Sequencing Markers. *Front. Plant Sci.* **2020**, *11*, 534. [CrossRef] [PubMed]
125. Ma, Y.; Coyne, C.J.; Grusak, M.A.; Mazourek, M.; Cheng, P.; Main, D.; McGee, R.J. Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). *BMC Plant Biol.* **2017**, *17*, 1–17. [CrossRef]
126. Qin, J.; Shi, A.; Mou, B.; Grusak, M.A.; Weng, Y.; Ravelombola, W.; Bhattarai, G.; Dong, L.; Yang, W. Genetic diversity and association mapping of mineral element concentrations in spinach leaves. *BMC Genomics* **2017**, *18*, 1–14. [CrossRef]
127. Tian, H.; Yang, Y.; Wang, R.; Fan, Y.; Yi, H.; Jiang, B.; Wang, L.; Ren, J.; Xu, L.; Zhang, Y.; et al. Screening of 200 core SNPs and the Construction of a Systematic SNP-DNA Standard Fingerprint Database with More Than 20,000 Maize Varieties. *Agriculture* **2021**, *11*, 597. [CrossRef]
128. Kuhn, D.N.; Groh, A.; Rahaman, J.; Freeman, B.; Arpaia, M.L.; Van den Berg, N.; Abeysekara, N.; Manosalva, P.; Chambers, A.H. Creation of an avocado unambiguous genotype SNP database for germplasm curation and as an aid to breeders. *Tree Genet. Genomes* **2019**, *15*, 71. [CrossRef]
129. Medina-lozano, I.; Bertolín, J.R.; Zufiaurre, R.; Diaz, A. Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.). *J. Vis. Exp.* **2020**, *160*, e61440. [CrossRef] [PubMed]
130. Dossou, S.S.K.; Xu, F.; You, J.; Zhou, R.; Li, D.; Wang, L. Widely targeted metabolome profiling of different colored sesame (*Sesamum indicum* L.) seeds provides new insight into their antioxidant activities. *Food Res. Int.* **2022**, *151*, 110850. [CrossRef] [PubMed]
131. Cheng, Z.Q.; Huang, X.Q.; Zhang, Y.Z.; Qian, J.; Yang, M.Z.; Wu, C.J.; Liu, J.F. Diversity in the content of some nutritional components in husked seeds of three wild rice species and rice varieties in Yunnan Province of China. *J. Integr. Plant Biol.* **2005**, *47*, 1260–1270. [CrossRef]
132. Esteras, C.; Rambla, J.L.; Sánchez, G.; López-Gresa, M.P.; González-Mas, M.C.; Fernández-Trujillo, J.P.; Bellés, J.M.; Granell, A.; Picó, M.B. Fruit flesh volatile and carotenoid profile analysis within the *Cucumis melo* L. species reveals unexploited variability for future genetic breeding. *J. Sci. Food Agric.* **2018**, *98*, 3915–3925. [CrossRef] [PubMed]
133. Burbano-Erazo, E.; León-Pacheco, R.I.; Cordero-Cordero, C.C.; López-Hernández, F.; Cortés, A.J.; Tofiño-Rivera, A.P. Multi-Environment Yield Components in Advanced Common Bean (*Phaseolus vulgaris* L.) × Tepary Bean (*P. acutifolius* A. Gray) Interspecific Lines for Heat and Drought tolerance. *Agronomy* **2021**, *11*, 1978. [CrossRef]
134. Herraiz, F.J.; Raigón, M.D.; Vilanova, S.; García-Martínez, M.D.; Gramazio, P.; Plazas, M.; Rodríguez-Burruezo, A.; Prohens, J. Fruit composition diversity in land races and modern pepino (*Solanum muricatum*) varieties and wild related species. *Food Chem.* **2016**, *203*, 49–58. [CrossRef] [PubMed]
135. Medina-Lozano, I.; Bertolín, J.R.; Díaz, A. Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: Vitamin C and anthocyanin content. *Food Chem.* **2021**, *359*. [CrossRef] [PubMed]
136. Li, K.; Wang, D.; Gong, L.; Lyu, Y.; Guo, H.; Chen, W.; Jin, C.; Liu, X.; Fang, C.; Luo, J. Comparative analysis of metabolome of rice seeds at three developmental stages using a recombinant inbred line population. *Plant J.* **2019**, *100*, 908–922. [CrossRef] [PubMed]
137. Jeon, J.E.; Kim, J.-G.; Fischer, C.R.; Mehta, N.; Dufour-Schroif, C.; Wemmer, K.; Mudgett, M.B.; Sattely, E. Pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* **2020**, *180*, 176–187. [CrossRef] [PubMed]
138. Raza, A. Metabolomics: A systems biology approach for enhancing heat stress tolerance in plants. *Plant Cell Rep.* **2020**. [CrossRef] [PubMed]
139. Jaganathan, D.; Bohra, A.; Thudi, M.; Varshney, R.K. Fine mapping and gene cloning in the post-NGS era: Advances and prospects. *Theor. Appl. Genet.* **2020**, *133*, 1791–1810. [CrossRef]
140. Sharma, V.; Gupta, P.; Priscilla, K.; Sharankumar; Hangargi, B.; Veershetty, A.; Ramrao, D.P.; Suresh, S.; Narasanna, R.; Naik, G.R.; et al. Metabolomics Intervention Towards Better Understanding of Plant Traits. *Cells* **2021**, *10*, 346. [CrossRef] [PubMed]
141. Bilbrey, E.A.; Williamson, K.; Hatzakis, E.; Miller, D.D.; Fresnedo-Ramírez, J.; Cooperstone, J.L. Integrating genomics and multiplatform metabolomics enables metabolite quantitative trait loci detection in breeding-relevant apple germplasm. *New Phytol.* **2021**, *232*, 1944–1958. [CrossRef] [PubMed]
142. Templer, S.E.; Ammon, A.; Pscheidt, D.; Ciobotea, O.; Schuy, C.; McCollum, C.; Sonnewald, U.; Hanemann, A.; Förster, J.; Ordon, F.; et al. Metabolite profiling of barley flag leaves under drought and combined heat and drought stress reveals metabolic QTLs for metabolites associated with antioxidant defense. *J. Exp. Bot.* **2017**, *68*, 1697–1713. [CrossRef] [PubMed]
143. Matros, A.; Houston, K.; Tucker, M.R.; Schreiber, M.; Berger, B.; Aubert, M.K.; Wilkinson, L.G.; Witzel, K.; Waugh, R.; Seiffert, U.; et al. Genome-wide association study reveals the genetic complexity of fructan accumulation patterns in barley grain. *J. Exp. Bot.* **2021**, *72*, 2383–2402. [CrossRef] [PubMed]
144. Mahalingam, R.; Sallam, A.H.; Steffenson, B.J.; Fiedler, J.D.; Walling, J.G. Genome-wide association analysis of natural variation in seed tocochromanols of barley. *Plant Genome* **2020**, *13*. [CrossRef] [PubMed]
145. Peng, M.; Shahzad, R.; Gul, A.; Subthain, H.; Shen, S.; Lei, L.; Zheng, Z.; Zhou, J.; Lu, D.; Wang, S.; et al. Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* **2017**, *8*, 1–12. [CrossRef] [PubMed]

146. Ferrão, L.F.V.; Johnson, T.S.; Benevenuto, J.; Edger, P.P.; Colquhoun, T.A.; Munoz, P.R. Genome-wide association of volatiles reveals candidate loci for blueberry flavor. *New Phytol.* **2020**, *226*, 1725–1737. [CrossRef] [PubMed]
147. Matros, A.; Liu, G.; Hartmann, A.; Jiang, Y.; Zhao, Y.; Wang, H.; Ebmeyer, E.; Korzun, V.; Schachschneider, R.; Kazman, E.; et al. Genome-metabolite associations revealed low heritability, high genetic complexity, and causal relations for leaf metabolites in winter wheat (*Triticum aestivum*). *J. Exp. Bot.* **2017**, *68*, 415–428. [CrossRef] [PubMed]
148. Wei, W.; Li, S.; Wang, Y.; Wang, B.; Fan, G.; Zeng, Q.; Zhao, F.; Xu, C.; Zhang, X.; Tang, T.; et al. Metabolome-Based Genome-Wide Association Study Provides Genetic Insights Into the Natural Variation of Foxtail Millet. *Front. Plant Sci.* **2021**, *12*, 665530. [CrossRef] [PubMed]
149. Zhang, W.; Alseekh, S.; Zhu, X.; Zhang, Q.; Fernie, A.R.; Kuang, H.; Wen, W. Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *Plant J.* **2020**, *104*, 613–630. [CrossRef]
150. Wang, Y. A draft genome, resequencing, and metabolomes reveal the genetic background and molecular basis of the nutritional and medicinal properties of loquat (*Eriobotrya japonica* (Thunb.) Lindl). *Hortic. Res.* **2021**, *8*, 231. [CrossRef] [PubMed]
151. Li, H.; Peng, Z.; Yang, X.; Wang, W.; Fu, J.; Wang, J.; Han, Y.; Chai, Y.; Guo, T.; Yang, N.; et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* **2013**, *45*, 43–50. [CrossRef] [PubMed]
152. Wen, W.; Li, D.; Li, X.; Gao, Y.; Li, W.; Li, H.; Liu, J.; Liu, H.; Chen, W.; Luo, J.; et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **2014**, *5*, 3438. [CrossRef] [PubMed]
153. Riedelsheimer, C.; Lisec, J.; Czedik-Eysenberg, A.; Sulpice, R.; Flis, A.; Grieder, C.; Altmann, T.; Stitt, M.; Willmitzer, L.; Melchinger, A.E. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8872–8877. [CrossRef] [PubMed]
154. Lipka, A.E.; Gore, M.A.; Magallanes-Lundback, M.; Mesberg, A.; Lin, H.; Tiede, T.; Chen, H.; Robin Buell, C.; Buckler, E.S.; Rocheford, T.; et al. Genome-Wide Association Study and Pathway-Level Analysis of Tocohromanol Levels in Maize Grain. *G3 Genes Genomes Genet.* **2013**, *3*, 1287–1299. [CrossRef]
155. Owens, B.F.; Gore, M.A.; Magallanes-Lundback, M.; Tiede, T.; Diepenbrock, C.H.; Kandianis, C.B.; Kim, E.; Cepela, J.; Mateos-Hernandez, M.; Robin Buell, C.; et al. A Foundation for Provitamin A Biofortification of Maize: Genome-Wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics* **2014**, *198*, 1699–1716. [CrossRef] [PubMed]
156. Liang, X.; Liu, S.; Wang, T.; Li, F.; Cheng, J.; Lai, J.; Qin, F.; Li, Z.; Wang, X.; Jiang, C. Metabolomics-driven gene mining and genetic improvement of tolerance to salt-induced osmotic stress in maize. *New Phytol.* **2021**, *230*, 2355–2370. [CrossRef]
157. Suwarno, W.B.; Pixley, K.V.; Palacios-Rojas, N.; Kaeppler, S.M.; Babu, R. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor. Appl. Genet.* **2015**, *128*, 851–864. [CrossRef] [PubMed]
158. Deng, M.; Li, D.; Luo, J.; Xiao, Y.; Liu, H.; Pan, Q.; Zhang, X.; Jin, M.; Zhao, M.; Yan, J. The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol. J.* **2017**, *15*, 1250–1263. [CrossRef]
159. Jin, M.; Zhang, X.; Zhao, M.; Deng, M.; Du, Y.; Zhou, Y.; Wang, S.; Tohge, T.; Fernie, A.R.; Willmitzer, L.; et al. Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis. *BMC Plant Biol.* **2017**, *17*, 17. [CrossRef]
160. Azmach, G.; Menkir, A.; Spillane, C.; Gedil, M. Genetic Loci Controlling Carotenoid Biosynthesis in Diverse Tropical Maize Lines. *G3 Genes Genomes Genet.* **2018**, *8*, 1049–1065. [CrossRef] [PubMed]
161. Wang, H.; Xu, S.; Fan, Y.; Liu, N.; Zhan, W.; Liu, H.; Xiao, Y.; Li, K.; Pan, Q.; Li, W.; et al. Beyond pathways: Genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. *Plant Biotechnol. J.* **2018**, *16*, 1464–1475. [CrossRef] [PubMed]
162. Wen, W.; Jin, M.; Li, K.; Liu, H.; Xiao, Y.; Zhao, M.; Alseekh, S.; Li, W.; de Abreu e Lima, F.; Brotman, Y.; et al. An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J.* **2018**, *93*, 1116–1128. [CrossRef] [PubMed]
163. Alves, M.L.; Bento-Silva, A.; Carbas, B.; Gaspar, D.; Paulo, M.; Brites, C.; Mendes-Moreira, P.; Brites, C.M.; Bronze, M.D.R.; Malosetti, M.; et al. Alleles to Enhance Antioxidant Content in Maize-A Genome-Wide Association Approach. *J. Agric. Food Chem.* **2020**, *68*, 4051–4061. [CrossRef]
164. Baseggio, M.; Murray, M.; Magallanes-Lundback, M.; Kaczmar, N.; Chamness, J.; Buckler, E.S.; Smith, M.E.; DellaPenna, D.; Tracy, W.F.; Gore, M.A. Genome-Wide Association and Genomic Prediction Models of Tocochromanols in Fresh Sweet Corn Kernels. *Plant Genome* **2019**, *12*, 180038. [CrossRef]
165. Diepenbrock, C.H.; Ilut, D.C.; Magallanes-Lundback, M.; Kandianis, C.B.; Lipka, A.E.; Bradbury, P.J.; Holland, J.B.; Hamilton, J.P.; Wooldridge, E.; Vaillancourt, B.; et al. Eleven biosynthetic genes explain the majority of natural variation in carotenoid levels in maize grain. *Plant Cell* **2021**, *33*, 882–900. [CrossRef]
166. Chatham, L.A.; Juvik, J.A. Linking anthocyanin diversity, hue, and genetics in purple corn. *G3 Genes Genomes Genet.* **2021**, *11*, jkaa062. [CrossRef]
167. Levina, A.V.; Hoekenga, O.; Gordin, M.; Broeckling, C.; De Jong, W.S. Genetic analysis of potato tuber metabolite composition: Genome-wide association studies applied to a nontargeted metabolome. *Crop. Sci.* **2021**, *61*, 591–603. [CrossRef]
168. Dong, X.; Gao, Y.; Chen, W.; Wang, W.; Gong, L.; Liu, X.; Luo, J. Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol. Plant* **2015**, *8*, 111–121. [CrossRef]
169. Brotman, Y.; Llorente-Wiegand, C.; Oyong, G.; Badoni, S.; Misra, G.; Anacleto, R.; Parween, S.; Pasion, E.; Tiozon, R.N.; Anonuevo, J.J.; et al. The genetics underlying metabolic signatures in a brown rice diversity panel and their vital role in human nutrition. *Plant J.* **2021**, *106*, 507–525. [CrossRef]

170. Chen, W.; Wang, W.; Peng, M.; Gong, L.; Gao, Y.; Wan, J.; Wang, S.; Shi, L.; Zhou, B.; Li, Z.; et al. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* **2016**, *7*, 12767. [CrossRef] [PubMed]
171. Matsuda, F.; Nakabayashi, R.; Yang, Z.; Okazaki, Y.; Yonemaru, J.I.; Ebana, K.; Yano, M.; Saito, K. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* **2015**, *81*, 13–23. [CrossRef] [PubMed]
172. Zhang, F.; Guo, H.; Huang, J.; Yang, C.; Li, Y.; Wang, X.; Qu, L.; Liu, X.; Luo, J. A UV-B-responsive glycosyltransferase, OsUGT706C2, modulates flavonoid metabolism in rice. *Sci. China Life Sci.* **2020**, *63*, 1037–1052. [CrossRef] [PubMed]
173. Li, X.; Tian, R.; Shao, Z.; Zhang, H.; Chu, J.; Li, W.; Kong, Y.; Du, H.; Zhang, C. Genetic loci and causal genes for seed fatty acids accumulation across multiple environments and genetic backgrounds in soybean. *Mol. Breed.* **2021**, *41*, 31. [CrossRef]
174. Wu, D.; Li, D.; Zhao, X.; Zhan, Y.; Teng, W.; Qiu, L.; Zheng, H.; Li, W.; Han, Y. Identification of a candidate gene associated with isoflavone content in soybean seeds using genome-wide association and linkage mapping. *Plant J.* **2020**, *104*, 950–963. [CrossRef] [PubMed]
175. Liu, J.Y.; Li, P.; Zhang, Y.W.; Zuo, J.F.; Li, G.; Han, X.; Dunwell, J.M.; Zhang, Y.M. Three-dimensional genetic networks among seed oil-related traits, metabolites and genes reveal the genetic foundations of oil synthesis in soybean. *Plant J.* **2020**, *103*, 1103–1124. [CrossRef] [PubMed]
176. Fang, K.; Xia, Z.; Li, H.; Jiang, X.; Qin, D.; Wang, Q.; Wang, Q.; Pan, C.; Li, B.; Wu, H. Genome-wide association analysis identified molecular markers associated with important tea flavor-related metabolites. *Hortic. Res.* **2021**, *8*, 42. [CrossRef]
177. Yamashita, H.; Uchida, T.; Tanaka, Y.; Katai, H.; Nagano, A.J.; Morita, A.; Ikka, T. Genomic predictions and genome-wide association studies based on RAD-seq of quality-related metabolites for the genomics-assisted breeding of tea plants. *Sci. Rep.* **2020**, *10*, 17480. [CrossRef]
178. Tieman, D.; Zhu, G.; Resende, M.F.R.; Lin, T.; Nguyen, C.; Bies, D.; Rambla, J.L.; Beltran, K.S.O.; Taylor, M.; Zhang, B.; et al. A chemical genetic roadmap to improved tomato flavor. *Plant Sci.* **2017**, *355*, 6323. [CrossRef]
179. Sauvage, C.; Segura, V.; Bauchet, G.; Stevens, R.; Do, P.T.; Nikoloski, Z.; Fernie, A.R.; Causse, M. Genome-Wide Association in Tomato Reveals 44 Candidate Loci for Fruit Metabolic Traits. *Plant Physiol.* **2014**, *165*, 1120–1132. [CrossRef]
180. Zhu, G.; Wang, S.; Huang, Z.; Zhang, S.; Liao, Q.; Zhang, C.; Lin, T.; Qin, M.; Peng, M.; Yang, C.; et al. Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **2018**, *172*, 249–261.e12. [CrossRef] [PubMed]
181. Burgos, E.; Belen De Luca, M.; Diouf, I.; de Haro, L.A.; Albert, E.; Sauvage, C.; Tao, Z.J.; Bermudez, L.; Asís, R.; Nesi, A.N.; et al. Validated MAGIC and GWAS population mapping reveals the link between vitamin E content and natural variation in chorismate metabolism in tomato. *Plant J.* **2021**, *105*, 907–923. [CrossRef] [PubMed]
182. Chen, J.; Hu, X.; Shi, T.; Yin, H.; Sun, D.; Hao, Y.; Xia, X.; Luo, J.; Fernie, A.R.; He, Z.; et al. Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol. J.* **2020**, *18*, 1722–1735. [CrossRef] [PubMed]
183. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef] [PubMed]
184. Duangjit, J.; Causse, M.; Sauvage, C. Efficiency of genomic selection for tomato fruit quality. *Mol. Breed.* **2016**, *36*, 29. [CrossRef]
185. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* **2016**, *9*, plantgenome2016-01. [CrossRef] [PubMed]
186. Dempewolf, H.; Baute, G.; Anderson, J.; Kilian, B.; Smith, C.; Guarino, L. Past and Future Use of Wild Relatives in Crop Breeding. *Crop. Sci.* **2017**, *57*, 1070–1082. [CrossRef]
187. Bouis, H.E.; Saltzman, A. Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. *Glob. Food Sec.* **2017**, *12*, 49–58. [CrossRef] [PubMed]
188. HarvestPlus. *Getting Biofortified Food on Everyone's Plate—2019 Annual Report*; HarvestPlus: Washington, DC, USA, 2019.
189. Sazawal, S.; Dhingra, U.; Dhingra, P.; Dutta, A.; Deb, S.; Kumar, J.; Devi, P.; Prakash, A. Efficacy of high zinc biofortified wheat in improvement of micronutrient status, and prevention of morbidity among preschool children and women—A double masked, randomized, controlled trial. *Nutr. J.* **2018**, *17*, 86. [CrossRef]
190. Palmer, A.C.; Healy, K.; Barffour, M.A.; Siamusantu, W.; Chileshe, J.; Schulze, K.J.; West, K.P.J.; Labrique, A.B. Provitamin A Carotenoid-Biofortified Maize Consumption Increases Pupillary Responsiveness among Zambian Children in a Randomized Controlled Trial. *J. Nutr.* **2016**, *146*, 2551–2558. [CrossRef]
191. Low, J.W.; Mwanga, R.O.M.; Andrade, M.; Carey, E.; Ball, A.-M. Tackling vitamin A deficiency with biofortified sweetpotato in sub-Saharan Africa. *Glob. Food Sec.* **2017**, *14*, 23–30. [CrossRef] [PubMed]
192. Prasanna, B.M.; Palacios-Rojas, N.; Hossain, F.; Muthusamy, V.; Menkir, A.; Dhliwayo, T.; Ndhlela, T.; San Vicente, F.; Nair, S.K.; Vivek, B.S.; et al. Molecular Breeding for Nutritionally Enriched Maize: Status and Prospects. *Front. Genet.* **2020**, *10*, 1392. [CrossRef] [PubMed]
193. Cruet-Burgos, C.; Cox, S.; Ioerger, B.P.; Perumal, R.; Hu, Z.; Herald, T.J.; Bean, S.R.; Rhodes, D.H. Advancing provitamin A biofortification in sorghum: Genome-wide association studies of grain carotenoids in global germplasm. *Plant Genome* **2020**, *13*, e20013. [CrossRef] [PubMed]
194. Izquierdo, P.; Astudillo, C.; Blair, M.W.; Iqbal, A.M.; Raatz, B.; Cichy, K.A. Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **2018**, *131*, 1645–1658. [CrossRef] [PubMed]

195. Krishnappa, G.; Rathan, N.D.; Sehgal, D.; Ahlawat, A.K.; Singh, S.K.; Singh, S.K.; Shukla, R.B.; Jaiswal, J.P.; Solanki, I.S.; Singh, G.P.; et al. Identification of Novel Genomic Regions for Biofortification Traits Using an SNP Marker-Enriched Linkage Map in Wheat (*Triticum aestivum* L.). *Front. Nutr.* **2021**, *8*, 669444. [CrossRef] [PubMed]
196. Garg, M.; Sharma, A.; Vats, S.; Tiwari, V.; Kumari, A.; Mishra, V.; Krishania, M. Vitamins in Cereals: A Critical Review of Content, Health Effects, Processing Losses, Bioaccessibility, Fortification, and Biofortification Strategies for Their Improvement. *Front. Nutr.* **2021**, *8*, 586815. [CrossRef] [PubMed]
197. Bhatt, V.; Muthysamy, V.; Jha, S.; Zunjare, R.U.; Baveja, A.; Sosad, S. Development of low phytic acid maize through marker assisted introgression of *lpa1-1* and *lpa2-1* genes. In Proceedings of the 13th Asian Maize Conference on and Expert Consultation on Maize for Food, Feed, Nutrition and Environmental Security, Ludhiana, India, 8–10 October 2018; pp. 143–144.
198. Storozhenko, S.; De Brouwer, V.; Volckaert, M.; Navarrete, O.; Blancquaert, D.; Zhang, G.F.; Lambert, W.; Van Der Straeten, D. Folate fortification of rice by metabolic engineering. *Nat. Biotechnol.* **2007**, *25*, 1277–1279. [CrossRef] [PubMed]
199. Díaz de La Garza, R.I.; Gregory, J.F.; Hanson, A.D. Folate biofortification of tomato fruit. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4218–4222. [CrossRef]
200. Bulley, S.; Wright, M.; Rommens, C.; Yan, H.; Rassam, M.; Lin-Wang, K.; Andre, C.; Brewster, D.; Karunairetnam, S.; Allan, A.C.; et al. Enhancing ascorbate in fruits and tubers through over-expression of the L-galactose pathway gene GDP-L-galactose phosphorylase. *Plant Biotechnol. J.* **2012**, *10*, 390–397. [CrossRef]
201. Pierce, E.C.; LaFayette, P.R.; Ortega, M.A.; Joyce, B.L.; Kopsell, D.A.; Parrott, W.A. Ketocarotenoid production in soybean seeds through metabolic engineering. *PLoS ONE* **2015**, *10*, e0138196. [CrossRef]
202. Holme, I.B.; Dionisio, G.; Brinch-Pedersen, H.; Wendt, T.; Madsen, C.K.; Vincze, E.; Holm, P.B. Cisgenic barley with improved phytase activity. *Plant Biotechnol. J.* **2012**, *10*, 237–247. [CrossRef] [PubMed]
203. Kamrani, M.; Kohnehrouz, B.B.; Gholizadeh, A. Cisgenic inhibition of the potato cold induced phosphorylase L gene expression and decrease in sugar contents. *Afr. J. Biotechnol.* **2011**, *10*, 10076–10082. [CrossRef]
204. De Vetten, N.; Wolters, A.M.; Raemakers, K.; Van der Meer, I.; Ter Stege, R.; Heeres, E.; Heeres, P.; Visser, R. A transformation method for obtaining marker-free plants of a cross-pollinating and vegetatively propagated crop. *Nat. Biotechnol.* **2003**, *21*, 439–442. [CrossRef] [PubMed]
205. Chawla, R.; Shakya, R.; Rommens, C.M. Tuber-specific silencing of *asparagine synthetase-1* reduces the acrylamide-forming potential of potatoes grown in the field without affecting tuber shape and yield. *Plant Biotechnol. J.* **2012**, *10*, 913–924. [CrossRef] [PubMed]
206. Gil-Humanes, J.; Pistón, F.; Hernando, A.; Alvarez, J.B.; Shewry, P.R.; Barro, F. Silencing of γ-gliadins by RNA interference (RNAi) in bread wheat. *J. Cereal Sci.* **2008**, *48*, 565–568. [CrossRef]
207. Brinch-Pedersen, H.; Borg, S.; Tauris, B.; Holm, P.B. Molecular genetic approaches to increasing mineral availability and vitamin content of cereals. *J. Cereal Sci.* **2007**, *46*, 308–326. [CrossRef]
208. Coca, M.; Peñas, G.; Gómez, J.; Campo, S.; Bortolotti, C.; Messeguer, J.; Segundo, B.S. Enhanced resistance to the rice blast fungus *Magnaporthe grisea* conferred by expression of a cecropin A gene in transgenic rice. *Planta* **2006**, *223*, 392–406. [CrossRef]
209. Girgi, M.; Breese, W.A.; Lörz, H.; Oldach, K.H. Rust and Downy Mildew Resistance in Pearl Millet (*Pennisetum glaucum*) Mediated by Heterologous Expression of the afp Gene from *Aspergillus giganteus*. *Transgenic Res.* **2006**, *15*, 313–324. [CrossRef]
210. Yadav, H.; Malik, K.; Kumar, S.; Jaiwal, P.K. Comparative regeneration in six bread wheat (*Triticum aestivum* L.) varieties from immature and mature scutella for developing efficient and genotype-independent protocol prerequisite for genetic improvement of wheat. *Vitr. Cell. Dev. Biol.-Plant* **2020**, *56*, 610–617. [CrossRef]
211. Sainger, M.; Chaudhary, D.; Dahiya, S.; Jaiwal, R.; Jaiwal, P.K. Development of an efficient in vitro plant regeneration system amenable to *Agrobacterium*-mediated transformation of a recalcitrant grain legume blackgram (*Vigna mungo* L. Hepper). *Physiol. Mol. Biol. Plants* **2015**, *21*, 505–517. [CrossRef]
212. Garg, M.; Sharma, N.; Sharma, S.; Kapoor, P.; Kumar, A.; Chunduri, V.; Arora, P. Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. *Front. Nutr.* **2018**, *5*, 12. [CrossRef] [PubMed]
213. Tang, G.; Qin, J.; Dolnikowski, G.G.; Russell, R.M.; Grusak, M.A. Golden Rice is an effective source of vitamin A. *Am. J. Clin. Nutr.* **2009**, *89*, 1776–1783. [CrossRef] [PubMed]
214. Smyth, S.J. Genetically modified crops, regulatory delays, and international trade. *Food Energy Secur.* **2017**, *6*, 78–86. [CrossRef]
215. Schouten, H.J.; Krens, F.A.; Jacobsen, E. Do cisgenic plants warrant less stringent oversight? *Nat. Biotechnol.* **2006**, *24*, 753. [CrossRef] [PubMed]
216. Rommens, C.M. All-native DNA transformation: A new approach to plant genetic engineering. *Trends Plant Sci.* **2004**, *9*, 457–464. [CrossRef] [PubMed]
217. Jiang, M.; Liu, Y.; Liu, Y.; Tan, Y.; Huang, J.; Shu, Q. Mutation of Inositol *1,3,4-trisphosphate 5/6-kinase6* Impairs Plant Growth and Phytic Acid Synthesis in Rice. *Plants* **2019**, *8*, 114. [CrossRef]
218. Sun, Y.; Jiao, G.; Liu, Z.; Zhang, X.; Li, J.; Guo, X.; Du, W.; Du, J.; Francis, F.; Zhao, Y.; et al. Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* **2017**, *8*, 298. [CrossRef]
219. Schouten, H.J.; Krens, F.A.; Jacobsen, E. Cisgenic plants are similar to traditionally bred plants: International regulations for genetically modified organisms should be altered to exempt cisgenesis. *EMBO Rep.* **2006**, *7*, 750–753. [CrossRef]

220. Zhu, S.; Duwal, A.; Su, Q.; Vossen, J.H.; Visser, R.G.F.; Jacobsen, E. Vector integration in triple *R* gene transformants and the clustered inheritance of resistance against potato late blight. *Transgenic Res.* **2013**, *22*, 315–325. [CrossRef]
221. Conner, A.J.; Barrell, P.J.; Baldwin, S.J.; Lokerse, A.S.; Cooper, P.A.; Erasmuson, A.K.; Nap, J.P.; Jacobs, J.M.E. Intragenic vectors for gene transfer without foreign DNA. *Euphytica* **2007**, *154*, 341–353. [CrossRef]
222. Schouten, H.J.; Jacobsen, E. Cisgenesis and intragenesis, sisters in innovative plant breeding. *Trends Plant Sci.* **2008**, *13*, 260–261. [CrossRef] [PubMed]
223. Bhatnagar, M.; Prasad, K.; Bhatnagar-Mathur, P.; Narasu, M.L.; Waliyar, F.; Sharma, K.K. An efficient method for the production of marker-free transgenic plants of peanut (*Arachis hypogaea* L.). *Plant Cell Rep.* **2010**, *29*, 495–502. [CrossRef] [PubMed]
224. Doshi, K.M.; Eudes, F.; Laroche, A.; Gaudet, D. Anthocyanin expression in marker free transgenic wheat and triticale embryos. *Vitr. Cell. Dev. Biol.-Plant* **2007**, *43*, 429–435. [CrossRef]
225. Vanblaere, T.; Szankowski, I.; Schaart, J.; Schouten, H.; Flachowsky, H.; Broggini, G.A.L.; Gessler, C. The development of a cisgenic apple plant. *J. Biotechnol.* **2011**, *154*, 304–311. [CrossRef]
226. Ling, F.; Zhou, F.; Chen, H.; Lin, Y. Development of marker-free insect-resistant indica rice by *Agrobacterium tumefaciens*-mediated co-transformation. *Front. Plant Sci.* **2016**, *7*, 1608. [CrossRef]
227. Delwaide, A.C.; Nalley, L.L.; Dixon, B.L.; Danforth, D.M.; Nayga, R.M.; Van Loo, E.J.; Verbeke, W. Revisiting GMOs: Are There Differences in European Consumers' Acceptance and Valuation for Cisgenically vs Transgenically Bred Rice? *PLoS ONE* **2015**, *10*, e0126060. [CrossRef]
228. Shew, A.M.; Nalley, L.L.; Danforth, D.M.; Dixon, B.L.; Nayga, R.M.; Delwaide, A.C.; Valent, B. Are all GMOs the same? Consumer acceptance of cisgenic rice in India. *Plant Biotechnol. J.* **2016**, *14*, 4–7. [CrossRef]
229. Edenbrandt, A.K.; Gamborg, C.; Thorsen, B.J. Consumers' Preferences for Bread: Transgenic, Cisgenic, Organic or Pesticide-free? *J. Agric. Econ.* **2018**, *69*, 121–141. [CrossRef]
230. Raj, R.S.; Singh, C.; Modi, A.; Subhash, N. Genetic transformation of lowland rice variety GR11 for drought tolerance and its ratification for upland paddy cultivation. *Indian J. Genet. Plant Breed.* **2015**, *75*, 30–40. [CrossRef]
231. Roberts, R.J. The Nobel Laureates' Campaign Supporting GMOs. *J. Innov. Knowl.* **2018**, *3*, 61–65. [CrossRef]
232. June, M.K. *International Service for the Acquisition of Agri-Biotech (ISAAA). Global Status of Commercialized Biotech/GM Crops: 2019*; ISAAA: Ithaca, NY, USA, 2020; ISBN 978-1-892456-69-9.
233. Purchase, I.F.H. What determines the acceptability of genetically modified food that can improve human nutrition? *Toxicol. Appl. Pharmacol.* **2005**, *207*, 19–27. [CrossRef] [PubMed]
234. Davison, J.; Ammann, K. New GMO regulations for old: Determining a new future for EU crop biotechnology. *GM Crop. Food* **2017**, *8*, 13–34. [CrossRef] [PubMed]
235. McDougall, P. *R&D Trends for Chemical Crop Protection Products and the Position of the European Market*; Phillips McDougall Ltd.: Pathhead, UK, 2013.
236. Holme, I.B.; Wendt, T.; Holm, P.B. Intragenesis and cisgenesis as alternatives to transgenic crop development. *Plant Biotechnol. J.* **2013**, *11*, 395–407. [CrossRef] [PubMed]
237. Hou, H.; Atlihan, N.; Lu, Z.X. New biotechnology enhances the application of cisgenesis in plant breeding. *Front. Plant Sci.* **2014**, *5*, 389. [CrossRef] [PubMed]
238. Russell, A.W.; Sparrow, R. The case for regulating intragenic GMOs. *J. Agric. Environ. Ethics* **2008**, *21*, 153–181. [CrossRef]
239. EFSA Panel on Genetically Modified Organisms (GMO). Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis. *EFSA J.* **2012**, *10*, 2561. [CrossRef]
240. European Food Safety Authority (EFSA). *International Scientific Workshop 'Risk Assessment Considerations for RNAi-Based GM Plants'*; EFSA Supporting Publication: Brussels, Belgium, 2014; Volume EN-705.
241. Environmental Protection Agency (EPA). *Pesticides; Data Requirements for Plant.-Incorporated Protectants (PIPs) and Certain Exemptions for PIPs*; EPA: Whasington, DC, USA, 2011; Volume 76.
242. De Steur, H.; Wesana, J.; Blancquaert, D.; Van Der Straeten, D.; Gellynck, X. The socioeconomics of genetically modified biofortified crops: A systematic review and meta-analysis. *Ann. N. Y. Acad. Sci.* **2017**, *1390*, 14–33. [CrossRef]
243. Xu, Y.; Li, P.; Zou, C.; Lu, Y.; Xie, C.; Zhang, X.; Prasanna, B.M.; Olsen, M.S. Enhancing genetic gain in the era of molecular breeding. *J. Exp. Bot.* **2017**, *68*, 2641–2666. [CrossRef]

ANNEX 2. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.mdpi.com/article/10.3390/ijms23063086/s1.

Supplemental File 1. Bibliographic search criteria to elaborate the present review.

This review was performed by a search of international literature in the databases PubMed, Web of Science and Scopus, with the following search terms in several fields within the different sections:

1. Introduction: ("biofortification") AND ("crop"); ("metabolomic") AND ("crop").

2. Exploring biodiversity: searching for outstanding material: ("SNP genotyping") AND ("crop").

3. Association between the traits of interest and the genomic regions: fishing for genes: ("GWAS") AND ("metabolomic") OR ("metabolic GWAS") OR ("mGWAS").

4. Introducing allelic variants to biofortify crops: ("breeding" OR transgen* OR cisgen* OR intragen*) AND ("biofortification")); (cisgen* AND intragen*).

5. Regulation of plant breeding methods: ("GMO" OR "transgenic") AND ("legislation")); ((cisgen* OR intragen*) AND ("legislation")).

The retrieved list of references was manually screened, on one hand, to be filtered and, on the other hand, to look for additional publications relevant to this review. No limitations of date, language, or study design were established. In addition, data from international organizations like FAO, UNICEF, EFSA, UK Government etc., were also consulted. The searches were conducted during 2021 and the beginning of 2022.

ANNEX 3

Medina-Lozano, I, Díaz, A (2021). Nutritional Value and Phytochemical Content of Crop Landraces and Traditional Varieties, in *Landraces – Traditional Variety and Natural Breed*, ed. A. Elkelish (IntechOpen), 95–116. ISBN 978-1-83968-718-1. https://doi.org/10.5772/intechopen.95514.

Chapter 6

Nutritional Value and Phytochemical Content of Crop Landraces and Traditional Varieties

Inés Medina-Lozano and Aurora Díaz

Abstract

Over the years, crops have been improved through breeding, mainly to increase production and, secondly, to introduce resistance to diseases and to achieve tolerance to abiotic stresses, these two latter by resorting to Crop Wild Relatives (CWR). This has resulted, in most cases, in homogeneous and nutritionally poor commercial varieties. Landraces and traditional varieties, barely taken into account, are key resources as they retain nutrients frequently "washed away" in the commercial varieties and also harbour a great genetic variability. They could represent a shortcut when compared to CWR in breeding, saving time and resources. The consumer's growing interest in health and food quality has caused breeders to redirect their attention toward them. This chapter provides information about the content in compounds with health benefits, such as phenolics, minerals, vitamins, etc., of landraces and traditional varieties of the most important crops, which could help to obtain healthier and more nutritious products.

Keywords: biofortification, carotenoids, micronutrients, health-promoting compounds, minerals, plant breeding, phenolic compounds, vitamins

1. Introduction

1.1 Landraces and traditional varieties: similarities, differences and comparison with wild species and commercial varieties

In the wide spectrum of plant material in terms of domestication and/or breeding, the concepts seem to be clear in both extremes, wild forms and commercial varieties. On one hand, the wild plants (either the Crop Wild Relatives, CWR, or those belonging to more distant gene pools) are those that have not been domesticated or subject to processes of artificial (human) selection and breeding. They do not exhibit traits typically present in cultivated plants, like uniform seed germination and homogeneous fruit ripening, or desirable characteristics present in those plants destined to human consumption, mainly related to quality (**Figure 1**). On the other hand, commercial varieties are those obtained by a breeding programme aimed to improve certain traits of the crop and that differ from other existing varieties by distinctive properties, which are uniformly expressed, and transferred in a stable way to the subsequent generations (**Figure 1**).

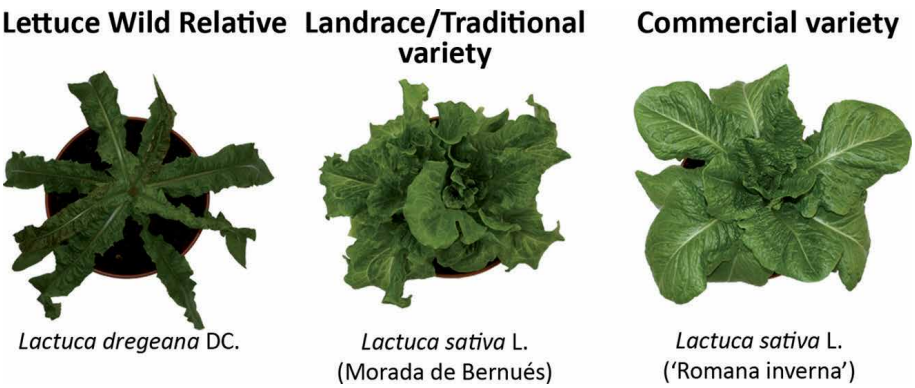


Figure 1.
Examples of phytogenetic resourses within the genus Lactuca: A lettuce wild relative (Lactuca dregeana DC.*) and two cultivated forms (Lactuca sativa* L.*), a landrace/traditional variety (Morada de Bernués), and a commercial variety ('Romana inverna').*

In between those two ends, a wide plethora of intermediate forms can be found. That is a grey area with blurred boundaries, what explains the general lack of consensus in even defining the plant material. In many cases, different terms have been used to refer to the same (or similar) type of plant, like ecotype, landrace, race, farmer variety, folk variety, local variety, traditional cultivar, etc. [1]. Even if some definitions are contradictory, there seems to be some recurrent ideas when authors refer to landraces and traditional varieties.

Landraces are profusely described in the literature as autochthonous cultivars or, at least, cultivars that have been grown in a certain area since ancestral times and, hence, are adapted to local growing conditions and uses through natural selection but without any active intervention from farmers. There are several terms difficult to verify in that definition. It does not seem easy to trace back the origin of the cultivars, especially if we take into account that the crop dispersals and the human migrations are inseparable. Besides, even if they have been cultivated in a region for a long period of time and, hence, they are adapted to the predominant environmental conditions, that does not imply that they exhibit a great tolerance to adverse conditions, biotic and abiotic stresses as stated before [1, 2]. Actually, the adverse edaphic, climatic and phytosanitary conditions would be mitigated even by the most traditional low input agricultural systems in comparison to those that the wild plants would have to face in the same region. Finally, it is difficult to defend the idea of farmers growing a cultivar for generations without carrying out any type of selection of the outstanding individuals, even if it is not fully conscious, as stated mainly in the earliest definitions [3, 4]. In fact, the agriculture procedures (seeding, harvesting ...) exert an artificial selection under which the most suitable genotypes for those cultural practises, prosper (and they probably rely on them for their survival, in return). Furthermore, in a scenario in which only the natural selection is acting, the resulting plants would probably be more similar to their wild relatives and less to the bred cultivars (and that is not the case with the landraces). In more recent definitions, the idea of a more or less directed human selection has been embraced [5–7], even if it cannot be considered a formal breeding programme [8].

In contrast with landraces, traditional varieties (also called folk varieties or farmer-bred varieties), have usually been defined like those that have been maintained by active selection and/or breeding by farmers. And, if this is the main difference between landraces and traditional varieties (as the latter are also cultivated locally and are well adapted to the particular climatic and growing conditions), is it really possible to distinguish them? How do we determine if a certain variety is the

product of merely natural selection or the human intervention has also played a role on it? Is it actually possible to separate both processes? It could be that the question is nowadays irrelevant and the important aspect is that, either if we call them landraces or traditional varieties (**Figure 1**), they consist of dynamic populations that harbour enough genetic variability to show adaptability to local conditions and plasticity to overcome eventual changes, even if they can be fairly uniform for the selected traits. That broad genetic base would explain that, under eventual adverse conditions, they are still able to yield stably (though moderately), as some genotypes within the population will possibly show a better performance. These aspects were early emphasised by the plant breeder Harlan [9] when stated that some of the most important characteristics of landraces are their genetic diversity and dynamism, what has also been adopted in more modern times by other authors [10, 11]. Harlan also pointed out that they are the result of millennia of natural and artificial selection, as a way of integrating these two indiscernible processes. Another approach to overcome this thorny aspect consisted of eliminating the type of selection undergone in the definition of the landraces [12]. Any realistic and updated definition of this type of plant material will have to include the impact of agriculture and, hence, the human influence in their evolution as proposed recently [13].

Another aspect that blurs the lines between landraces and traditional varieties is the gene flow between them. With the availability of molecular markers and Next-Generation Sequencing (NGS) techniques, it is possible to trace the allele introgression from cultivated (all types) to wild plants and *vice versa*. Even if there were landraces exclusively product of natural selection and traditional varieties obtained by men selection, obviously, gene transfer could have also happened between them, especially taking into account that exchanging plant material is a common practise among farmers.

In any case, the main differences when compared to commercial varieties are that landraces and traditional varieties do not always have a traceable origin, they exhibit a great diversity and, precisely for that reason, most of their traits are less uniform within them and less stable through the descendants.

1.2 Importance and conservation of landraces and traditional varieties in germplasm banks worldwide

The great variability harboured by the landraces and traditional varieties is one of their most outstanding characteristics. Historically, all this richness had been preserved and used (a vicious circle of cause and effect) by the agriculturalists. That situation started to change when the erosion of the plant genetic resources became patent for scientists and breeders, not only in the case of landraces and traditional varieties, but also concerning the wild species. Since then, the germplasm banks have assumed a principal role in safeguarding this plant biodiversity [14]. The strategy has revealed itself so successful that, according to the World Information and Early Warning System (WIEWS) on Plant Genetic Resources for Food and Agriculture (PGRFA), approximately 5.4 million accessions are being conserved in over 710 genebanks from 103 countries and 17 international/regional centres [15]. Landraces and traditional varieties represent the heart of the collections, what becomes obvious when the numbers of the different types of plant resources are consulted. As an example, in Genesys, which is a portal that supplies not only seeds, but also characterisation and evaluation data about PGRFA from germplasm banks around the world [16], landraces and traditional varieties account for the highest proportion of accessions (37%), followed by breeding and research material (27%), advanced and improved cultivars (19%), and finally, wild forms (17%) (**Figure 2**).

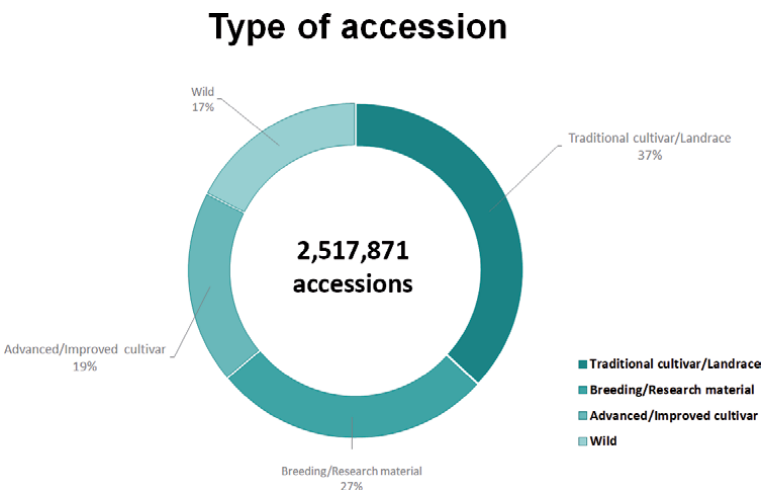


Figure 2.
Relative amount of the different types of accessions attending to their biological status (excluding the "not specified" material) hold at Genesys [16], the online platform which harbour information about PGRFA conserved in genebanks worldwide.

The high genetic variability exhibited by landraces and traditional varieties obviously translates into characteristics that could be desirable in modern varieties, particularly those related to their nutritional value and content of health-promoting compounds, which is the subject under discussion in this chapter. In modern breeding programmes, flavour selection has prevailed over nutritional quality. That explains why, for instance, modern lettuce varieties have almost lost their ancestral bitterness. That is a direct consequence of the drastic decrease in the content of sesquiterpene lactones, which are not only responsible for the bitter taste but have also beneficial properties for the plant itself and for the animals that feed on it [17]. In other cases, the main objective has not been to eliminate flavours detrimental to the taste but to enhance the pleasant ones. This is the case for sweet corn. Its sugar content has been escalating over the last decades by the selection of varieties with an increasing polysaccharide content: sugar-enhanced varieties, supersweet or extrasweet varieties, high sugar varieties... [18]. The side effect has been the disappearance of non-sweet dark-grain primitive varieties rich in anthocyanins, which happen to be powerful antioxidants with an important role for health by preventing cardiovascular diseases and by presenting anti-cancer activity [19, 20]. The landraces and traditional varieties were shaped under very different criteria, what does not necessary implies that they are better, for instance, from a nutritional perspective, than any commercial variety, though they contribute to increase the agrodiversity and to enrich the diet. In this sense, the germplasm banks can act as gene reservoir to improve crops, allowing us to dive for valuable characteristics to obtain all types of plant material (coming from crosses between different traditional varieties, between traditional varieties and CWR, between traditional varieties and breeding material ...), using both conventional and biotechnological tools.

2. Essential micronutrients

Essential micronutrients are nutrients that must be obtained in the diet as they cannot be synthesised by the human body. They are required in small quantities and usually consist of vitamins and minerals. Micronutrients play vital roles in human

health, so their deficiencies can be devastating. These deficiencies, also known as hidden hunger, are mainly consequence of micronutrient low concentrations in the daily diet, resulting in malnutrition that is considered an important global problem of public health, especially in developing countries. In addition, the impact is more serious in women of reproductive age (especially pregnant women) and under-five children due to their higher micronutrient requirements. In fact, maternal and child malnutrition or micronutrient deficiencies affect approximately half of the world's population [21]. Nevertheless, hidden hunger also affects developed countries due to low quality food or bad habits, like extreme diets to lose weight or alcohol and drug abuse.

Generally, fruits and vegetables are the sources of vitamins and minerals, but their concentrations in most plant foods are not sufficient to reach the daily dietary requirements, even if the recommended amounts are consumed [22]. Besides, micronutrient content usually depends on the plant genotype, among other factors like environmental conditions, time of harvest, etc. Cases in which landraces and traditional varieties of important crops exhibit higher contents of micronutrients than commercial and modern cultivars are described here. They actually could play a key role in human health by supplying an enhanced nutrition.

2.1 Organic micronutrients: vitamins

Vitamins are a diverse group of organic molecules that are essential in trace quantities for a proper metabolism of all living organisms and are mainly synthesised by plants and microorganisms. Vitamins can be classified into fat-soluble (A, D, E and K) or water-soluble (vitamin B-complex, C and H) compounds. Their main function is to act as cofactors for many enzymes and as natural antioxidants, both in plants and animals. In addition, some vitamins play specific roles, for example, in human vision (vitamin A) [23] or as hormones implied in calcium and phosphorus homeostasis in the blood stream (vitamin D) [24], and many of them are indispensable to prevent chronic diseases [19, 20].

Plants, mostly fruits and vegetables, are the main source of vitamins for humans. However, their concentration in the edible portions of most important crops is usually below the recommended daily intake, which entails important implications for global human health [24]. Interestingly, some landraces exceed these minimal requirements or, at least, they are richer than commercial cultivars in these micronutrients, especially for vitamins A, C and E.

2.1.1 Vitamin A

Vitamin A is a fat-soluble vitamin group that includes retinol and its derivatives, like retinoic acid and retinal, among many others [25]. Besides, among the large group of compounds known as carotenoids, there are some that can act as precursors of vitamin A, known generically as provitamin A, such as α-carotene, β-cryptoxanthin and β-carotene, the most abundant and nutritionally active within them all. The richest sources of vitamin A are from animal origin, whereas carotenoids are synthesised mainly by plants, but also by some fungi and microorganisms.

Carotenoids play important roles in plant metabolism: acting as pigments in different tissues, mediating plant–animal interaction for pollination or seed dispersal, participating in cell photoprotection against photooxidative damage and heat stress, and protecting membranes from lipid peroxidation thanks to their antioxidant activity [26].

In humans, provitamin A is involved in vision, immune responses, cellular growth, development and reproduction [23]. Vitamin A deficiency is one of the micronutrient deficiencies with more devastating consequences for health. It is the main cause of preventable blindness in children and pregnant women, especially in low-income countries, and raises the risk of suffering several diseases and of dying as a result of severe infections. Between 250,000 and 500,000 vitamin A deficient children become blind every year, half of them dying 12 months later [27]. Therefore, it is a question of the utmost importance to know what plant-based foods contain high levels of provitamin A.

The β-carotene content was measured in two Spanish landraces of tomato (*Solanum lycopersicum* L.) and in the commercial variety 'Moneymaker' [28]. A higher concentration of this carotenoid was found in green fruits of the two landraces when compared to 'Moneymaker', whereas in ripe fruits, only the landrace Negro Yeste had a higher amount, even more than double. Also in comparison with the commercial variety 'Moneymaker', three tomato landraces, two from Italy and one from Guatemala, showed a significantly higher β-carotene content [29]. In other study carried out in melon (*Cucumis melo* L.), landraces of different origins exhibited the highest levels of β-cryptoxanthin and β-carotene compared with commercial melons [30]. In an analysis of the β-carotene content of mungbean (*Vigna radiata* L. Wilzeck), the landrace VI000323 B-G happened to have grains significantly richer than two improved mungbean lines at their maturity stage [31]. Though modern wheat (*Triticum* spp.) varieties were not analysed, old varieties (from the 1900–1960 breeding period) were included as reference, and the average values obtained for β-carotene and β-cryptoxanthin were significantly higher in the wheat landraces than in the old cultivars [32]. Also in landraces of pepino (*Solanum muricatum* Ait.) from the Andean region [33] and in the landrace G-4615 of sweet potato (*Ipomoea batatas* (L.) Lam.) from Solomon Islands [34], higher contents of β-carotene than in improved varieties have been obtained.

2.1.2 Vitamin C

Vitamin C is a water-soluble vitamin that comprises ascorbic acid (AA), the main biologically active form, and its oxidation product, dehydroascorbic acid (DHAA), easily convertible into AA in the human body [35]. In plants, vitamin C plays relevant roles in metabolic and defence processes, as it is an important antioxidant in the ascorbate-glutathione pathway, it protects enzymes with prosthetic metal ions, it is a cofactor for many enzymes (including those involved in cell wall synthesis), it is involved in photosynthesis and respiration, etc. [36].

In humans, it is crucial in some metabolic processes as it participates in collagen formation and inorganic iron absorption, and contributes to a healthy state by reducing the cholesterol levels, preventing chronic diseases and enhancing the immune system by its antioxidant action [37]. The main consequence of vitamin C deficiency is scurvy and, although relatively few people suffer this deficiency, the benefits of the micronutrient are evident, so it is important to find vitamin C-rich plant food.

Some studies have reported a higher content in vitamin C in crop landraces with respect to commercial varieties. For example, 17 Spanish melon traditional varieties were evaluated and most of them had significantly higher values of AA when compared with 10 commercial accessions of reference, in some cases even doubling the AA values of the commercial variety within the same market class (Piel de Sapo, Yellow, Ananás...) [38]. Traditional varieties of lettuce (*Lactuca sativa* L.) from Aragón (Spain) have also been reported to have higher average contents in vitamin C than commercial varieties, especially AA content [39, 40]. Some Spanish

landraces of eggplant (*Solanum melongena* L.) had also a higher concentration of both, AA and DHAA, than commercial hybrids [41]. In other experiment, four to seven traditional varieties of tomato contained higher concentrations of vitamin C than the commercial variety 'Baghera', with significant differences for the traditional varieties CIDA-62 and BGW-004123. In addition, CIDA-62 fruits showed the highest antioxidant activity, whereas the lowest was observed in the commercial variety [42]. Other authors also reported 10 indeterminate tomato landraces that exhibited significantly higher AA contents than the commercial variety 'Moneymaker' [29]. In analyses of the AA content in accessions of garlic (*Allium sativum* L.) from Plugia region (Italy), the six landraces evaluated had a higher content than the commercial cultivar used as reference [43]. Higher contents of total vitamin C have also been obtained in grains of the mungbean landrace VI000323 B-G from Taiwan [31], in the Greek onion (*Allium cepa* L.) landrace Vatikiotiko [44] and in two rare landraces of Italian turnip (*Brassica rapa* L. subspecies *rapa*) [45] when compared with commercial and improved varieties.

2.1.3 Vitamin E

Vitamin E is a fat-soluble vitamin group made up of tocopherols and tocotrienols, a group of lipid-soluble compounds. Both tocopherols and tocotrienols can present four different methylated forms, α, β, γ and δ, and although all of them are antioxidants, α-tocopherol is the most abundant form and has the highest activity [46].

In plants, the main function of vitamin E is as antioxidant, quenching and scavenging singlet oxygen, controlling the extent of lipid peroxidation, preserving the integrity of the membranes, and protecting against photoinhibition and photo-oxidative stresses [36].

In humans, vitamin E also acts as a potent antioxidant and it is involved in multiple physiological processes, such as regulation of gene expression and cognitive performance. Besides, vitamin E plays a key role in maintaining a healthy state by controlling the inflammation, enhancing the immune function and preventing light-induced pathologies of the skin and eyes, and degenerative disorders like cardiovascular diseases, atherosclerosis and cancer. Its deficiency is common in developing countries and affects mainly children and the elderly, and can cause haemolytic anaemia in premature babies and neurological and ophthalmological disorders as well as myopathy in children. In developed countries it is rare and only appears in some stages of development, such as in premature babies, and specific conditions, like in digestive and genetic pathologies [24].

A total of 28 Korean accessions of soybean (*Glycine max* L.) were evaluated and the highest total tocopherol contents were observed in the seeds of the 7 landraces analysed, especially in HaNagari, in comparison with the modern cultivars developed by cross-breeding, in which paradoxically the content decreased gradually with the year of registration [47]. Furthermore, a strong negative correlation between tocopherol contents and lipid peroxidation was observed (what demonstrates the vitamin E role in oxidative stress tolerance), with the soybean landraces showing the lowest lipid peroxidation. In wheat, higher contents of tocopherols and tocotrienols were obtained for some landraces in comparison with modern cultivars when individual genotypes were analysed [48]. Hazelnut (*Corylus avellana* L.) is also a good source of vitamin E and an Argentinian landrace has been reported to have the highest total tocopherol content in comparison with different commercial cultivars [49]. The total contents of tocopherols and tocotrienols, as well as total vitamin E, were higher in traditional red rice (*Oryza sativa* L.) varieties than in three light brown new-improved varieties [50].

2.2 Mineral micronutrients

Mineral micronutrients are inorganic elements required in small quantities to play vital functions in both, plants and animals. The nutrient classifications are dynamic and, sometimes, even contradictory. That is because, on one hand, the limit between small and big quantities that determine the inclusion of an element in the micronutrient or macronutrient list can result arbitrary. On the other hand, new discoveries about the participation of some elements in important physiological mechanisms cause their transfer from the "nonessential" to the "essential" list. Magnesium (Mg) is a clear example of discrepancies on the first criteria as, depending on the author, is described as micronutrient or macronutrient as ranks in an intermediate position in terms of recommended daily allowances [51]. Regarding the second criteria, some minerals like boron (B) have been known to be essential for plant nutrition for a long time but it has not been until a few decades ago that its important effect on human nutrition was noted [52].

In plants, mineral micronutrients participate in different physiological processes of primary and secondary metabolism, like photosynthesis, electron transfer, activation of enzymes, cell defence, hormone perception, gene regulation... So, mineral deficiencies affect the plant life cycle seriously, causing even plant death in the severest cases [53].

In humans, more than 22 mineral elements (altogether micro- and macronutrients) are essential and they can be obtained with an appropriate diet [51]. Nevertheless, mineral deficiencies are very common, especially in developing countries, and their consequences, such as learning disabilities in children, increased morbidity and mortality, low productivity at work..., are detrimental for humans. Iron (Fe), zinc (Zn) and iodine (I) are the mineral elements most frequently lacking in the diet and their deficiencies, together with vitamin A deficiency, are responsible for about 12% of the deaths among under-five children globally [21]. Fe is important for oxygen transport and haemoglobin formation, and its deficiency is the main cause of preventable iron-deficiency anaemia, poor cognitive development, and maternal and child deaths [54, 55]. Zn plays a central role in growth, development and in the normal functioning of the immune system, so its deficiency hampers growth, alters immunity and also causes diarrhoea among children [56, 57]. Moreover, both deficiencies are also associated with childhood stunting. I is a component of the thyroid hormones and a strong antioxidant. Its deficiency can also cause growth impairments and, in addition, thyroid enlargement (goitre), hypothyroidism, pregnancy loss, infant mortality and cognitive and neuron psychological impairments [58]. On the other hand, manganese (Mn), copper (Cu) or selenium (Se) deficiencies are not a global issue, but they are common in some populations of developing countries, specifically in parts of China, India and Africa [51].

Many landraces of horticultural crops are reported to present higher contents of minerals and oligoelements than commercial varieties. In a study carried out with seeds of Turkish lentils (*Lens culinaris* Medik.), the average values of all the micro-minerals quantified (Cu, Fe, Mn, and Zn) were higher in the landraces than in the commercial cultivars, being Kahmar1 the richest in Zn and Cu, Diykub in Fe, and Kahmar2 in Mn [22]. Also in Turkey, higher contents of Zn and Se have been observed in common bean (*Phaseoulus vulgaris* L.) landraces than in modern varieties, especially in the landrace LR05 [59]. The Greek onion landrace Vatikiotiko [44] and a Greek garlic landrace [60] were both richer in Zn, Mn and Fe than well-established onion cultivars and hybrids commercialised in Greece and a garlic commercial cultivar used as control, respectively. In addition, the mineral content of the onion landrace was even higher than the suggested by USDA (United States Department of Agriculture) for raw onions as a standard reference, especially for

Fe. Results obtained in chickpea (*Cicer arietinum* L.) revealed that landraces from Kyrgyzstan presented higher average values for Fe, Mn, Cu and Zn compared with a breeding line [61]. In Andean landraces of pepino [33], in several eggplant landraces from Spain and Cuba [62], and in landraces of mungbean [31], higher contents in Fe and Zn than in commercial and modern varieties have also been reported.

For cereal crops there are also several studies in which landraces are reported to be richer in mineral micronutrients than commercial varieties. Wheat is one of the most important cereal crops worldwide and there are many studies on wheat landraces. The maximum contents in Fe, Cu, Zn, Mn, and Se were observed in wheat landraces from Canary Islands in comparison with the commercial cultivar 'Vitrón' [63]. Other authors [64] also reported landraces and old cultivars of wheat with a higher average concentration of B, Cu, Fe, and Zn, and of Cu, Fe, Zn, and Mn, respectively, than modern cultivars. Similarly, the average content in Fe, Zn, Mn, Cu, and strontium (Sr) in wheat grain was reported to be significantly higher in 12 Sicilian landraces than in 3 modern cultivars [65]. Other study found seven Afghan wheat landraces with higher content in Fe than reference lines in three different locations [66]. In the case of rice, two Indian landraces showed a higher content in Zn in brown and even polished (considered a poor source of micronutrients) grains than the commercial variety 'BPT 5204' ('Samba Mahsuri'), very appreciated for its high yield and quality [67].

3. Health-promoting phytochemicals

Plant-based foods are rich in different phytochemicals with health-promoting properties for the human body, in spite of not being essential nutrients. Polyphenols and carotenoids are the most important ones among these plant phytochemicals. Unlike micronutrients, their deficiencies in humans are not devastating, but their health benefits are very significant.

3.1 Polyphenols

Phenolic compounds (monophenols and polyphenols) are one of the most abundant and widely distributed groups of chemicals in plants, with more than 8,000 structurally-different compounds currently identified [68]. Particularly, polyphenols are characterised by the presence of aromatic rings with one or more hydroxyl groups and, depending on the basic chemical structure, they are classified in at least 10 different types. However, there is a growing tendency to group them in 2 main categories: flavonoid (flavones, flavonols, flavanols, flavanones, isoflavones, and anthocyanins) and non-flavonoid (phenolic acids, stilbenes, lignans, xanthones, and tannins) compounds. In plants, polyphenols are involved, on one hand, in crucial biological processes, such as cell division, development, hormonal regulation, reproduction, photosynthesis, pigmentation and pollinator attraction, and, on the other hand, in protection mechanisms against oxidative damage due to radiation or biotic stress (pathogens), among other causes, thanks to their antioxidant properties [69].

Polyphenols seems to be the main contributor to the total antioxidant activities of fruits, with flavonoids being the most abundant in human diets. The health-promoting effects associated with phenolic compounds include the elimination of free radicals, as well as the prevention of chronic diseases, such as cancer, diabetes and cardiovascular and neurodegenerative diseases [68].

There are a number of studies in which different polyphenols are more abundant in horticultural crop landraces than in commercial cultivars. This could be because

some polyphenols contribute to the bitterness and astringency of the food, what could have been negatively selected in modern breeding programmes. Tomato is one of the most important crops worldwide and it is very rich in polyphenols. Several Italian and Spanish landraces have been reported to have higher contents of total phenolic compounds than the commercial varieties 'Brigade' and 'Moneymaker', with significant higher levels of the flavonoids quercetin-3-rutinoside, kaempferol-*O*-rutinoside and kaempferol-*O*-glucoside in the case of the Spanish landraces [29, 70]. Nevertheless, polyphenols are abundant in many other crops. For example, different Spanish landraces of eggplant exhibited the highest average and individual contents of total phenolic compounds when compared with several commercial cultivars in two independent studies [41, 62]. Other study found higher levels of chlorogenic acid in three Italian landraces of carrots (*Daucus carota* L.) in comparison with a commercial cultivar used as reference [71]. Landraces of pepino from the Andean region also exhibited a higher average content of total phenolics than commercial cultivars [33]. Two rare Italian landraces of turnip showed similar concentration of total phenols between them, which was up to a 61% higher than in the commercial genotype also included in the study [45]. An Ecuadorian landrace of sweet potato showed the highest content in two particular anthocyanins (peonidin and cyaniding glucosides) when compared with several improved varieties [34]. Regarding phenolic acids and flavonoids, significant higher contents were observed in landraces of mungbean [9], garlic [43], and apple (*Malus domestica* Borkh.) [72], in comparison with improved lines and commercial varieties. Finally, in winery by-products from Majorcan landraces of grape (*Vitis vinifera* L.), the highest values of total anthocyanins, tannins, and total phenolic compounds were observed in the Escursac red landrace, with the commercial variety 'Cabernet Sauvignon' used as reference [73].

In the case of cereals, also some landraces have been reported to be richer in polyphenols than commercial cultivars. In extracts of wheat bread flour, the landrace Biancola showed higher contents of flavonoids and total phenolic compounds than three modern cultivars, as well as higher reducing power and lipid peroxidation inhibition levels [74]. Similarly, the landrace Gentil Rosso had a much higher amount of total, free, and bound polyphenols than three modern and five old cultivars [75]. In extracts of wheat grains, the highest contents of the 13 phenolic compounds identified were found in landraces when compared with commercial cultivars, especially in Tumminia SG3, Tripolino, Scavuzza, and Urria [76]. In maize (*Zea mays* L.), several Mexican landraces have been reported to have the highest content of phenylpropanoids in comparison with two commercial genotypes, especially Sinaloa 35, which contained exceptionally high levels of diferulates [77]. Also in maize, the Italian landrace Rostato Rosso contained a higher concentration of anthocyanins than an inbred line and a hybrid assayed [78]. Finally, in rice, traditional red-grained varieties of Sri Lanka exhibited significantly stronger antioxidant activity and higher total phenolic content in both, bran and grains, than light brown-grained newly improved varieties, with proanthocyanidins and phenolic acids among the most abundant phenolic compounds identified [50].

3.2 Carotenoids

Carotenoids are the second most abundant natural pigments, behind only chlorophyll, with more than 750 different structures known until now. They are synthesised by photosynthetic organisms (bacteria, algae and plants) and by some non-photosynthetic bacteria and fungi. They can be classified in two main groups: carotenes, composed of carbon and hydrogen atoms, such as α-carotene, β-carotene, and lycopene, among others; and xanthophylls, that are oxygenated hydrocarbon derivatives,

like lutein, cryptoxanthin, violaxanthin, zeaxanthin, etc. [79]. Carotenoids play key roles in several biological processes in plants. Apart from some of them being vitamin A precursors (as mentioned above), they are also precursors of the plant hormones abscisic acid (ABA) and strigolactones (SLs), they are one of the most important attractants to pollinators thanks to their pigmentation and fragrances (provided by volatile carotenoids), and they also participate in development, photosynthesis, photomorphogenesis and photoprotection processes [26].

The antioxidant potential of carotenoids is very important in human health due to their ability to reduce and, sometimes, prevent the development of various ROS (reactive oxygen species)-mediated disorders, such as cardiovascular diseases, cancer and neurological and photosensitive pathologies [80]. As humans are not able to synthesise these compounds, it is interesting to find crops rich in carotenoids. Vitamin A precursors (α-carotene, β-carotene and β-cryptoxanthin) have been described previously, so they are not dealt with here. Lycopene is the carotenoid responsible for tomato's red colour and it has been reported to be more abundant in two Spanish traditional varieties of tomato than in the commercial variety 'Baghera' [42]. In addition, one of these traditional varieties showed the strongest antioxidant activity. In two other studies carried out in tomato, not only lycopene, but also lutein content were significantly higher in a Spanish landrace and in three Italian landraces, respectively, than in the commercial variety 'Moneymaker' [28, 29]. Higher levels of lutein were also found in three Italian landraces of carrot, especially in the Tiggiano Yellow-Purple landrace [71], and in the melon landrace Casca de Carvalho [30] in comparison with commercial varieties. Cereal grains are also rich in carotenoids, especially lutein and zeaxanthin [81]. In this sense, several landraces of wheat exhibited higher levels of both compounds than old cultivars used as reference [32]. Finally, higher contents of lutein were also found in kernels of some maize traditional varieties from Italy, especially in Storo, in comparison to the hybrid B73/MO17, used as control [82].

4. Applications

As we all know, malnutrition is a public health problem with global dimensions. In 2019, almost 690 million people, 8.9% of the world population, were undernourished, mostly in developing countries. Beside this, about 2 billion people in the world suffered moderate or severe food insecurity, i.e. they did not have regular access to safe, nutritious, and sufficient food that year [83]. Overweight is also a growing matter of concern. In addition, since Green Revolution, the main objective of crop improvement programmes has been yield increase, what has resulted in a nutrient decrease in foodstuffs, contributing to malnutrition. However, quality has started to receive higher priority and agriculture objectives are undergoing changes from yield gains to the production of nutrient-rich food crops in sufficient amounts.

A search for crop landraces and traditional varieties with an enhanced nutritional value could be an interesting approach to combat nutrient deficiencies because, as seeing above, some of them are richer in micronutrients and health-promoting phytochemicals. However, they do not always cover minimal nutrient requirements and they are usually adapted to local environmental conditions. Therefore, a more feasible measure could be developing nutritionally enhanced foods with an increased bioavailability of nutrients for the human population. These efforts are normally directed toward raising the levels of minerals, vitamins, amino acids, and antioxidant compounds, as well as improving fatty acid composition in the edible portion of crop plants [84]. Crops with a higher nutritional value

can be obtained by agronomic practices, conventional plant breeding, and modern biotechnological techniques.

4.1 Fortification

Fortification through agronomic practices or traditional fortification consists of the physical addition of micronutrients to the plants to improve their nutritional quality. It is generally achieved by using mineral fertilisers to increase their content, bioavailability and/or transport from the soil to the edible portion of the plant. Plant growth-promoting soil microorganisms can also be used [85]. This approach is simple and fast but requires regular applications in every crop season, what can increases costs, and also needs supervision in order not to reach toxicity levels, both in the environment and for humans.

One example of this approach is the Se fortification through foliar application in different wheat genotypes [86]. The greater Se accumulations were obtained in the grains of the landrace Timilia and the obsolete variety 'Cappelli' when compared with modern varieties, with an increase of up to 35-fold in mineral grain concentration at the maximum Se application. In another study, fortification with I was carried out in the carrot landrace Carota di Polignano through foliar fertilisation in open field experiments and through both, foliar fertilisation and fertigation with nutrient solution, in greenhouse experiments [87]. In open field, the root content in I increased a 51% and a 194% with low and high levels of the fertiliser, respectively, when compared with untreated carrots, whereas in greenhouse, the I content increased a 9% and only with the fertigation.

4.2 Biofortification

Quite the opposite that the fortification, the biofortification consists of developing crops with a higher nutritional value *per se*, either through conventional breeding or through genetic engineering, without the need of external micronutrient addition. That means that the plants are able to synthesise greater amounts of the particular micronutrients.

Biofortification is a one-time investment and offers a long-term and cost-effective approach to prevent malnutrition: once a crop has been biofortified, no more costs, like adding fertilisers to the soil or fortificants to the processed food are needed. In addition, low-income countries could develop biofortified crops through traditional practices, so in theory, low cultivation and production costs are feasible [88]. Reducing the amount of fertilisers required to obtain a more nutritious crop has also unarguable environmental benefits. Nevertheless, biofortification is not the final solution but an additional tool to combat malnutrition.

4.2.1 Biofortification through conventional plant breeding

Biofortification through conventional plant breeding requires crosses between parent lines rich in nutrients and recipient lines that present desirable agronomic traits during several generations. This is a time-consuming method, though sustainable. However, this conventional biofortification relies on genetic variability, which is usually limited in commercial cultivar gene pools, especially of staple crops. Landraces and traditional varieties are an adequate alternative here, thanks to their wide genetic diversity. This approach has been applied to a wide variety of crops, especially since HarvestPlus Challenge Programme was launched in 2003 to develop biofortified staple food crops with enhanced essential micronutrients through plant conventional breeding [89].

Technique	Crop	Landrace or traditional variety	Enhanced trait	Method	Achievement	Reference
Agronomic practices	Wheat	Landrace Timilia; obsolete variety 'Capelli'	Se	Foliar fertilisation	↑ [Se] (up to 35-fold)	[86]
	Carrot	Landrace Carota di Polignano	I	Foliar fertilisation	↑ 51% and 194% with low and high levels of fertiliser, respectively	[87]
				Fertigation with nutrient solution	↑ 9%	
Conventional plant breeding	Rice	Traditional variety Zawa Bonday	Fe	Modern variety ('IR72') × traditional variety	Improved line with ↑ [Fe] (about 21 ppm in brown rice)	[90]
	Rice	Landrace Chittimuthyalu	Zn	Modern variety ('IR64') × landrace	Hybrid with ↑ [Zn] (26.9 mg/kg)	[91]
	Maize	Landrace ITA0370005	Carotenoids	Single cross: landrace × landrace (same population)	Hybrid with a ↑ [carotenoid] already commercialised	[92]
	Tomato	Landrace San Marzano	Polyphenols, tannins, flavonoids	Multiple crosses: landrace × landrace (same population)	Hybrid ('Torpedino di Fondi') with ↑ [polyphenols] and ↑ antioxidant activity in pink ripeness stage	[93]
	Eggplant	Nine landraces from Spain (ANS24, ANS26, ANS6, IVIA25, IVIA371, IVIA400, IVIA604, MUS8, VS22, VS9), one from China (ASIS1), and one from Cuba (SUDS5)	Polyphenols, Fe, Zn	Multiple crosses: landrace × landrace (different landraces)	Collection of hybrids with ↑ [phenolic compounds], ↑ [Fe], and ↑ [Zn]	[62]
	Eggplant	Landrace Almagro	Reduced prickliness	Backcrosses: three non-prickly commercial varieties × landrace	Improved pure line (H15) with nutritional properties of Almagro and ↓ prickliness	[94]
Modern biotechnology	Rice	Landrace Krabe	Seed yield	CRISPR-Cas9	Mutants with Krabe nutritional propierties and ↑ seed yield	[95]

Table 1.
Fortified and biofortified crops through different approaches by using landraces and traditional varieties.

Nevertheless, there is not a large number of studies carried out in landraces (**Table 1**). For example, in the International Rice Research Institute (IRRI) programme, an improved line (IR68144-3B-2-2-3) with a high concentration of Fe in the grain was obtained through a cross between a high-yield variety ('IR72') and a traditional variety (Zawa Bonday) from India. This new variety was reported to have about 80% more Fe than the commercial variety 'IR64' [90]. Useful information have been collected about the Zn content of different mapping populations of rice including wild germplasm, landraces and varieties, as well as hybrids [91]. Using 'IR64' as one of the parents, the hybrid with the highest Zn content (26.9 mg/kg) resulted from a cross with the landrace Chittimuthyalu. A collection of 14 hybrids between different landraces of eggplant has also been characterised [62]. These hybrids exhibited a higher average content of phenolics, as well as Fe and Zn, than commercial varieties. Zn average concentration was also higher in the hybrids than in the landraces tested. A maize hybrid with a high carotenoid content has also been identified [92]. It is a single-cross hybrid developed from the landrace ITA0370005 and it is currently being used by an Italian beer brewer. The metabolite profile and the antioxidant activity of the tomato hybrid Torpedino di Fondi (TF), developed from the landrace San Marzano (SM), has been characterised in two ripening stages, pink and red, both considered ideal for fresh consumption. In comparison with SM, pink TF tomatoes exhibited the highest content of total polyphenols, tannins, and flavonoids besides the greatest antioxidant activity [93]. Within a breeding programme, the eggplant landrace Almagro, known to contain higher values of vitamin C and total phenolics than regular varieties, but also having higher prickle presence, was used as recurrent parent in a backcross, whereas three non-prickly eggplant accessions were used as donors of this desirable trait [94]. Finally, an improved pure line (H15) with the Almagro eggplant ideotype and reduced prickliness was developed.

4.2.2 Biofortification through modern biotechnological techniques

Biofortification can be tackled through the genetic transformation of crops to express desirable genes from a plant species, independently of their taxonomic status, or even from other type of organisms, in the plant of interest to increase their nutrient content and bioavailability. This approach overcomes the limitation of the availability of genetic variability, allows the transfer of several genes simultaneously, and makes possible to biofortify crops with particular nutrients that are not naturally produced by themselves. Biofortification through transgenesis implies large investment of time, resources and researching: it is necessary to identify and characterise gene functions previously, and then, use these genes to transform crops. However, once the crop has been biofortified, it becomes a cost-effective approach [96].

The cisgenesis is a very interesting alternative to the transgenesis. With this approach, only genetic material from either the same species, or close relatives that hybridise naturally with it, is introduced [97]. In this way, the pool of genes available is exactly the same than when classical breeding methods are used. Cisgenic crops are subject to the same regulation than transgenic crops, but the EFSA (European Food Safety Authority) have concluded that cisgenics pose similar risks than plants obtained by conventional breeding [98]. Furthermore, the consumer's acceptance of cisgenics is greater than of transgenics [99].

Furthermore, the application of modern biotechnological techniques to landraces also allows the development of crops with higher yield, as it has been achieved recently [95]. The CRISPR-Cas9 technique was applied to the African rice landrace Kabre, considered a valuable resource, obtaining mutants with significantly improved seed yield and low lodging by disrupting genes known to control seed size and/or yield (**Table 1**).

5. Conclusion

In spite of not having been widely used in fortification and biofortification, especially with modern biotechnological approaches, crop landraces and traditional varieties could be key to improve the nutritional quality of food crops, as they can provide the desired genetic variability without sexual incompatibility barriers to overcome. Hopefully, in the near future there could be less restrictive regulations about the use of these biotechnological tools in crop breeding.

Acknowledgements

This work was funded by the projects RTA2017-00093-00-00 from the National Institute for Agricultural and Food Research and Technology (INIA) and LMP164_18 from the Government of Aragón; and by the Operational Programme FEDER Aragón 2014-2020 and the European Social Fund from the European Union [Grupo Consolidado A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genética"]. We gratefully acknowledge the Vegetable Germplasm Bank of Zaragoza (BGHZ-CITA, Spain) for supplying the seeds used for this work. I. M.-L. was granted with a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Spanish State Research Agency (AEI).

Conflict of interest

The authors declare no conflict of interest.

Author details

Inés Medina-Lozano[1,2] and Aurora Díaz[1,2]*

1 Department of Horticulture, Agrifood Research and Technology Centre of Aragon (CITA), Zaragoza, Spain

2 AgriFood Institute of Aragon – IA2 (CITA-University of Zaragoza), Zaragoza, Spain

*Address all correspondence to: adiazb@cita-aragon.es

References

[1] Zeven, A.C. Landraces: A review of definitions and classifications. Euphytica. 1998;104:127-139. DOI: 10.1023/A:1018683119237

[2] Mansholt, U.J. Van Pesch Plantenteelt, beknopte handleiding tot de kennis van den Nederlandschen landbouw. In Plantenteelt; 3rd ed. Zwolle; 1909. 228 p.

[3] von Rümker, K. Die systematische Einteilung und Benennung der Getreidesorten für praktische Zwecke. Jahrb. der Dtsch. landwirtschafts-Gesellschaft. 1908;23:137-167

[4] Fruwirth, C.; Roemer, T. Einführung in die landwirtschaftlichen Pflanzenzüchtung. Berlin; 1921. 150 p.

[5] Bellon, M.R.; Brush, S.B. Keepers of maize in Chiapas, Mexico. Econ. Bot. 1994;48:196-209. DOI: 10.1007/BF02908218

[6] Prospéri, J.; Demarquet, F.; Angevain, M.; Mansat, P. Évaluation agronomique de variétés de pays de sainfoin (*Onobrychis sativa* L.) originaires du sud-est de la France. Agronomie. 1994;14:285-298. DOI: 10.1051/agro:19940502

[7] Louette, D.; Charrier, A.; Berthaud, J. *In situ* conservation of maize in Mexico: Genetic diversity and maize seed management in a traditional community. Econ. Bot. 1997;51:20-38. DOI: 10.1007/BF02910401

[8] Teshome, A.; Baum, B.R.; Fahrig, L.; Torrance, J.K.; Arnason, T.J.; Lambert, J.D. Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in North Shewa and South Welo, Ethiopia. Euphytica. 1997;97:255-263. DOI: 10.1023/A:1003074008785

[9] Harlan, J.R. Our vanishing genetic resources. Science. 1975;188:618-621. DOI: 10.1126/science.188.4188.617

[10] Villa, T.C.C.; Maxted, N.; Scholten, M.; Ford-Lloyd, B. Defining and identifying crop landraces. Plant Genet. Resour. 2005;3:373-384. DOI: 10.1079/pgr200591

[11] Del Greco, A.; Negri, V.; Maxted, N. Report of a task force on on-farm conservation and management. In Proceedings of the Second Meeting; Stegelitz, Germany: Rome: Biodiversity International; 2007; p. 19-20

[12] Negri, V.; Maxted, N.; Veteläinen, M. European landrace conservation: an introduction. In Veteläinen, M., Negri, V., Maxted, N., editors. European Landraces: On-farm Conservation, Management and Use: Biodiversity Technical Bulletin No. 15. Rome, Italy: Biodiversity International; p. 275-282

[13] Casañas, F.; Simó, J.; Casals, J.; Prohens, J. Toward an evolved concept of landrace. Front. Plant Sci. 2017;8:145. DOI: 10.3389/fpls.2017.00145

[14] Mallor, C.; Díaz, A. Melon germplasm characteristics, diversity, preservation and uses. In Walton M, editors. Germplasm: Characteristics, Diversity and Preservation. New York: Nova Science Publishers; 2016. p. 1-26

[15] FAO. Food and Agriculture Organization of United Nations [Internet]. 2020. Available from: http://www.fao.org/ [Accessed: 2020-11-17]

[16] Genesys. Gateway to genetic resources [Internet]. 2020. Available from: https://www.genesys-pgr.org/a/overview [Accessed: 2020-11-19]

[17] Chadwick, M.; Trewin, H.; Gawthrop, F.; Wagstaff, C. Sesquiterpenoids lactones: Benefits to plants and people. Int. J. Mol. Sci. 2013;14:12780-12805. DOI: 10.3390/ijms140612780

[18] Lertrat, K.; Pulam, T. Breeding for increased sweetness in sweet corn. Int. J. Plant Breed. 2007;1:27-30

[19] He, J.; Monica Giusti, M. Anthocyanins: Natural colorants with health-promoting properties. Annu. Rev. Food Sci. Technol. 2010;1:163-187. DOI: 10.1146/annurev.food.080708.100754

[20] Yousuf, B.; Gul, K.; Wani, A.A.; Singh, P. Health benefits of anthocyanins and their encapsulation for potential use in food systems: A review. Crit. Rev. Food Sci. Nutr. 2016;56:2223-2230. DOI: 10.1080/10408398.2013.805316

[21] Ahmed, T.; Hossain, M.; Sanin, K.I. Global burden of maternal and child undernutrition and micronutrient deficiencies. Ann. Nutr. Metab. 2012;61:8-17. DOI: 10.1159/000345165

[22] Karaköy, T.; Erdem, H.; Baloch, F.S.; Toklu, F.; Eker, S.; Kilian, B.; Özkan, H. Diversity of macro-and micronutrients in the seeds of lentil landraces. Sci. World J. 2012;2012:710412. DOI: 10.1100/2012/710412

[23] Haskell, M.J.; Brown, K.H. Maternal vitamin A vutriture and the vitamin A content of human milk. J. Mammary Gland Biol. Neoplasia. 1999;4:243-257. DOI: 10.1023/A:1018745812512

[24] Fitzpatrick, T.B.; Basset, G.J.C.; Borel, P.; Carrari, F.; DellaPenna, D.; Fraser, P.D.; Hellmann, H.; Osorio, S.; Rothan, C.; Valpuesta, V.; Caris-Veyrat, C. Fernie, A.R. Vitamin deficiencies in humans: Can plant science help? Plant Cell. 2012;24:395-414. DOI: 10.1105/tpc.111.093120

[25] Olson, J.A. Vitamin A. In Decker M, editors. Handbook of Vitamins. New York: Eastern Hemisphere Distribution; 2001. p. 1-50

[26] DellaPenna, D.; Pogson, B.J. Vitamin synthesis in plants: Tocopherols and carotenoids. Annu. Rev. Plant Biol. 2006;57:711-738. DOI: 10.1146/annurev.arplant.56.032604.144301

[27] WHO. World Health Organization [Internet]. 2020. Available from: https://www.who.int/nutrition/topics/vad/en/ [Accessed: 2020-10-30]

[28] Massaretto, I.L.; Albaladejo, I.; Purgatto, E.; Flores, F.B.; Plasencia, F.; Egea-Fernández, J.M.; Bolarin, M.C.; Egea, I. Recovering tomato landraces to simultaneously improve fruit yield and nutritional quality against salt stress. Front. Plant Sci. 2018;871:1778. DOI: 10.3389/fpls.2018.01778

[29] Scarano, A.; Olivieri, F.; Gerardi, C.; Liso, M.; Chiesa, M.; Chieppa, M.; Frusciante, L.; Barone, A.; Santino, A.; Rigano, M.M. Selection of tomato landraces with high fruit yield and nutritional quality under elevated temperatures. J. Sci. Food Agric. 2020;100:2791-2799. DOI: 10.1002/jsfa.10312

[30] Esteras, C.; Rambla, J.L.; Sánchez, G.; López-Gresa, M.P.; González-Mas, M.C.; Fernández-Trujillo, J.P.; Bellés, J.M.; Granell, A.; Picó, M.B. Fruit flesh volatile and carotenoid profile analysis within the *Cucumis melo* L. species reveals unexploited variability for future genetic breeding. J. Sci. Food Agric. 2018;98:3915-3925. DOI: 10.1002/jsfa.8909

[31] Ebert, A.W.; Chang, C.H.; Yan, M.R.; Yang, R.Y. Nutritional composition of mungbean and soybean sprouts compared to their adult growth stage. Food Chem. 2017;237:15-22. DOI: 10.1016/j.foodchem.2017.05.073

[32] Hussain, A.; Larsson, H.; Kuktaite, R.; Olsson, M.E.; Johansson, E. Carotenoid content in organically produced wheat: Relevance for human nutritional health on consumption. Int. J. Environ. Res. Public Health. 2015;12:14068-14083. DOI: 10.3390/ijerph121114068

[33] Herraiz, F.J.; Raigón, M.D.; Vilanova, S.; García-Martínez, M.D.; Gramazio, P.; Plazas, M.; Rodríguez-Burruezo, A.; Prohens, J. Fruit composition diversity in land races and modern pepino (*Solanum muricatum*) varieties and wild related species. Food Chem. 2016;203:49-58. DOI: 10.1016/j.foodchem.2016.02.035

[34] Drapal, M.; Rossel, G.; Heider, B.; Fraser, P.D. Metabolic diversity in sweet potato (*Ipomoea batatas*, Lam.) leaves and storage roots. Hortic. Res. 2019;6:2. DOI: 10.1038/s41438-018-0075-5

[35] Lee, S.K.; Kader, A.A. Preharvest and postharvest factors influencing vitamin C content of horticultural crops. Postharvest Biol. Technol. 2000;20:207-220. DOI: 10.1016/S0925-5214(00)00133-2

[36] Ishikawa, T.; Dowdle, J.; Smirnoff, N. Progress in manipulating ascorbic acid biosynthesis and accumulation in plants. Physiol. Plant. 2006;126:343-355. DOI: 10.1111/j.1399-3054.2006.00640.x

[37] Carr, A.C.; Frei, B. Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. Am. J. Clin. Nutr. 1999;69:1086-1107. DOI: 10.1093/ajcn/69.6.1086

[38] Escribano, S.; Lázaro, A. Physicochemical and nutritional evaluation of Spanish melon landraces. Plant Genet. Resour. 2017;15:177-186. DOI: 10.1017/S1479262115000507

[39] Medina-Lozano, I.; Bertolín, J.R.; Zufiaurre, R.; Diaz, A. Improved UPLC-UV method for the quantification of vitamin C in lettuce varieties (*Lactuca sativa* L.) and crop wild relatives (*Lactuca* spp.). J. Vis. Exp. 2020;160:e61440. DOI: 10.3791/61440

[40] Medina-Lozano, I.; Bertolín, J.R.; Díaz, A. (in press). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. Food Chem.

[41] San José, R.; Sánchez-Mata, M.-C.; Cámara, M.; Prohens, J. Eggplant fruit composition as affected by the cultivation environment and genetic constitution. J. Sci. Food Agric. 2014;94:2774-2784. DOI: 10.1002/jsfa.6623

[42] Gonzalez-Cebrino, F.; Lozano, M.; Ayuso, M.C.; Bernalte, M.J.; Vidal-Aragon, M.C.; Gonzalez-Gomez, D. Characterization of traditional tomato varieties grown in organic conditions. Spanish J. Agric. Res. 2011;9:444-452. DOI: 10.5424/sjar/20110902-153-10

[43] Bonasia, A.; Conversa, G.; Lazzizera, C.; Loizzo, P.; Gambacorta, G.; Elia, A. Evaluation of garlic landraces from Foggia province (Puglia region; Italy). Foods. 2020;9:850. DOI: 10.3390/foods9070850

[44] Petropoulos, S.A.; Fernandes, Â.; Barros, L.; Ferreira, I.C.F.R.; Ntatsi, G. Morphological, nutritional and chemical description of "Vatikiotiko", an onion local landrace from Greece. Food Chem. 2015;182:156-163. DOI: 10.1016/j.foodchem.2015.03.002

[45] Conversa, G.; Lazzizera, C.; Bonasia, A.; Rotonda, P. La; Elia, A. Nutritional characterization of two rare landraces of turnip (*Brassica rapa*. var. *rapa*) tops and their on-farm conservation in Foggia province. Sustainability. 2020;12:3842. DOI: 10.3390/su12093842

[46] Fryer, M.J. The antioxidant effects of thylakoid Vitamin E (α-tocopherol). Plant. Cell Environ. 1992;15:381-392. DOI: 10.1111/j.1365-3040.1992.tb00988.x

[47] Lee, Y.Y.; Park, H.M.; Hwang, T.Y.; Kim, S.L.; Kim, M.J.; Lee, S.K.; Seo, M.J.; Kim, K.J.; Kwon, Y.U.; Lee,

S.C.; Kim, Y.H. A correlation between tocopherol content and antioxidant activity in seeds and germinating seeds of soybean cultivars. J. Sci. Food Agric. 2014;95:819-827. DOI: 10.1002/jsfa.6963

[48] Hussain, A.; Larsson, H.; Olsson, M.E.; Kuktaite, R.; Grausgruber, H.; Johansson, E. Is organically produced wheat a source of tocopherols and tocotrienols for health food? Food Chem. 2012;132:1789-1795. DOI: 10.1016/j.foodchem.2011.11.141

[49] Cittadini, M.C.; Martín, D.; Gallo, S.; Fuente, G.; Bodoira, R.; Martínez, M.; Maestri, D. Evaluation of hazelnut and walnut oil chemical traits from conventional cultivars and native genetic resources in a non-traditional crop environment from Argentina. Eur. Food Res. Technol. 2020;246:833-843. DOI: 10.1007/s00217-020-03453-8

[50] Gunaratne, A.; Wu, K.; Li, D.; Bentota, A.; Corke, H.; Cai, Y.Z. Antioxidant activity and nutritional quality of traditional red-grained rice varieties containing proanthocyanidins. Food Chem. 2013;138:1153-1161. DOI: 10.1016/j.foodchem.2012.11.129

[51] White, P.J.; Broadley, M.R. Biofortifying crops with essential mineral elements. Trends Plant Sci. 2005;10:586-593. DOI: 10.1016/j.tplants.2005.10.001

[52] Bolt, H.M.; Duydu, Y.; Başaran, N.; Golka, K. Boron and its compounds: current biological research activities. Arch. Toxicol. 2017;91:2719-2722. DOI: 10.1007/s00204-017-2010-1

[53] Vatansever, R.; Ozyigit, I.I.; Filiz, E. Essential and beneficial trace elements in plants, and their transport in roots: a review. Appl. Biochem. Biotechnol. 2016;181:464-482. DOI: 10.1007/s12010-016-2224-3

[54] Subramaniam, G.; Girish, M. Iron deficiency anemia in children. Indian J. Pediatr. 2015;82:558-564. DOI: 10.1007/s12098-014-1643-9

[55] Abbaspour, N.; Hurrell, R.; Kelishadi, R. Review on iron and its importance for human health. J. Res. Med. Sci. 2014;19:164-174

[56] Prasad, A.S. Discovery of human zinc deficiency: Its impact on human health and disease. Adv. Nutr. 2013;4:176-190. DOI: 10.3945/an.112.003210.176

[57] Roohani, N.; Hurrell, R.; Kelishadi, R.; Schulin, R. Zinc and its importance for human health: An integrative review. J. Res. Med. Sci. 2013;18:144-157. DOI: 10.1016/j.foodpol.2013.06.008

[58] Zimmermann, M.B.; Jooste, P.L.; Pandav, C.S. Iodine-deficiency disorders. Lancet. 2008;372:1251-1262. DOI: 10.1016/S0140-6736(08)61005-3

[59] Celmeli, T.; Sari, H.; Canci, H.; Sari, D.; Adak, A.; Eker, T.; Toker, C. The nutritional content of common bean (*Phaseolus vulgaris* L.) landraces in comparison to modern varieties. Agronomy. 2018;8:166. DOI: 10.3390/agronomy8090166

[60] Petropoulos, S.A.; Fernandes, Â.; Ntatsi, G.; Petrotos, K.; Barros, L.; Ferreira, I.C.F.R. Nutritional value, chemical characterization and bulb morphology of Greek garlic landraces. Molecules. 2018;23:319. DOI: 10.3390/molecules23020319

[61] Torutaeva, E.; Asanaliev, A.; Prieto-Linde, M.L.; Zborowska, A.; Ortiz, R.; Bryngelsson, T.; Garkava-Gustavsson, L. Evaluation of microsatellite-based genetic diversity, protein and mineral content in chickpea accessions grown in Kyrgyzstan. Hereditas. 2014;151:81-90. DOI: 10.1111/hrd2.00042

[62] Raigón, M.D.; Prohens, J.; Muñoz-Falcón, J.E.; Nuez, F. Comparison of eggplant landraces and commercial

varieties for fruit content of phenolics, minerals, dry matter and protein. J. Food Compos. Anal. 2008;21:370-376. DOI: 10.1016/j.jfca.2008.03.006

[63] Hernández Rodríguez, L.; Afonso Morales, D.; Rodríguez Rodríguez, E.; Díaz Romero, C. Minerals and trace elements in a collection of wheat landraces from the Canary Islands. J. Food Compos. Anal. 2011;24:1081-1090. DOI: 10.1016/j.jfca.2011.04.016

[64] Hussain, A.; Larsson, H.; Kuktaite, R.; Johansson, E. Mineral composition of organically grown wheat genotypes: Contribution to daily minerals intake. Int. J. Environ. Res. Public Health. 2010;7:3442-3456. DOI: 10.3390/ijerph7093442

[65] Sciacca, F.; Allegra, M.; Licciardello, S.; Roccuzzo, G.; Torrisi, B.; Virzì, N.; Brambilla, M.; Romano, E.; Palumbo, M. Potential use of Sicilian landraces in biofortification of modern durum wheat varieties: evaluation of caryopsis micronutrient concentrations. Cereal Res. Commun. 2018;46:124-134. DOI: 10.1556/0806.45.2017.056

[66] Kondou, Y.; Manickavelu, A.; Komatsu, K.; Arifi, M.; Kawashima, M.; Ishii, T.; Hattori, T.; Iwata, H.; Tsujimoto, H.; Ban, T.; Matsui, M. Analysis of grain elements and identification of best genotypes for Fe and P in Afghan wheat landraces. Breed. Sci. 2016;66:676-682. DOI: 10.1270/jsbbs.16041

[67] Neeraja, C.N.; Kulkarni, K.S.; Babu, P.M.; Rao, D.S.; Surekha, K.; Babu, V.R. Transporter genes identified in landraces associated with high zinc in polished rice through panicle transcriptome for biofortification. PLoS One. 2018;13:e0192362. DOI: 10.1371/journal.pone.0192362

[68] Lima, G.P.P.; Vianello, F.; Corrêa, C.R.; Campos, R.A. da S.; Borguini, M.G. Polyphenols in fruits and vegetables and its effect on human health. Food Nutr. Sci. 2014;5:1065-1082. DOI: 10.4236/fns.2014.511117

[69] Parr, A.J.; Bolwell, G.P. Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. J. Sci. Food Agric. 2000;80:985-1012. DOI: 10.1002/(sici)1097-0010(20000515)80:7<985::aid-jsfa572>3.3.co;2-z

[70] Siracusa, L.; Patanè, C.; Avola, G.; Ruberto, G. Polyphenols as chemotaxonomic markers in Italian "long-storage" tomato genotypes. J. Agric. Food Chem. 2011;60:309-314. DOI: 10.1021/jf203858y

[71] Scarano, A.; Gerardi, C.; D'Amico, L.; Accogli, R.; Santino, A. Phytochemical analysis and antioxidant properties in colored Tiggiano carrots. Agriculture. 2018;8:102. DOI: 10.3390/agriculture8070102

[72] Jakobek, L.; Barron, A.R. Ancient apple varieties from Croatia as a source of bioactive polyphenolic compounds. J. Food Compos. Anal. 2016;45:9-15. DOI: 10.1016/j.jfca.2015.09.007

[73] Garau, M.C.; González-Centeno, M.R.; Luna, J.M.; Negre, A.; Rosselló, C.; Femenia, A. Potential of landrace winery byproducts (*Vitis vinifera* L.) as a source of phenolic compounds with antioxidant properties. J. Int. des Sci. la Vigne du Vin. 2015;49:241-252. DOI: 10.20870/oeno-one.2015.49.4.45

[74] Falcinelli, B.; Calzuola, I.; Gigliarelli, L.; Torricelli, R.; Polegri, L.; Vizioli, V.; Benincasa, P.; Marsili, V. Phenolic content and antioxidant activity of wholegrain breads from modern and old wheat (*Triticum aestivum* L.) cultivars and ancestors enriched with wheat sprout powder. Ital. J. Agron. 2018;13:297-302. DOI: 10.4081/ija.2018.1220

[75] Migliorini, P.; Spagnolo, S.; Torri, L.; Arnoulet, M.; Lazzerini, G.; Ceccarelli, S. Agronomic and quality characteristics of old, modern and mixture wheat varieties and landraces for organic bread chain in diverse environments of northern Italy. Eur. J. Agron. 2016;79:131-141. DOI: 10.1016/j.eja.2016.05.011

[76] Bianco, M. Lo; Siracusa, L.; Dattilo, S.; Venora, G.; Ruberto, G. Phenolic fingerprint of sicilian modern cultivars and durum wheat landraces: A tool to assess biodiversity. Cereal Chem. 2017;94:1045-1051. DOI: 10.1094/CCHEM-06-17-0125-R

[77] Bily, A.C.; Burt, A.J.; Ramputh, A.I.; Livesey, J.; Regnault-Roger, C.; Philogène, B.R.; Arnason, J.T. HPLC-PAD-APCI assay of phenylpropanoids in cereals. Phytochem. Anal. 2004;15:9-15. DOI: 10.1002/pca.735

[78] Bernardi, J.; Stagnati, L.; Lucini, L.; Rocchetti, G.; Lanubile, A.; Cortellini, C.; De Poli, G.; Busconi, M.; Marocco, A. Phenolic profile and susceptibility to fusarium infection of pigmented maize cultivars. Front. Plant Sci. 2018;9:1189. DOI: 10.3389/fpls.2018.01189

[79] Nisar, N.; Li, L.; Lu, S.; Khin, N.C.; Pogson, B.J. Carotenoid metabolism in plants. Mol. Plant. 2015;8:68-82. DOI: 10.1016/j.molp.2014.12.007

[80] Fiedor, J.; Burda, K. Potential role of carotenoids as antioxidants in human health and disease. Nutrients. 2014;6:466-488. DOI: 10.3390/nu6020466

[81] Panfili, G.; Fratianni, A.; Irano, M. Improved normal-phase high-performance liquid chromatography procedure for the determination of carotenoids in cereals. J. Agric. Food Chem. 2004;52:6373-6377. DOI: 10.1021/jf0402025

[82] Puglisi, D.; Landoni, M.; Cassani, E.; Toschi, I.; Lucchini, G.; Cesari, V.; Borlini, G.; Pilu, R. Traditional farmers' varieties: a valuable source of genetic variability for biofortification programs. Maydica. 2018;63:1-10

[83] FAO; IFAD; UNICEF; WFP; WHO. The state of food security and nutrition in the world 2020. Transforming food systems for affordable healthy diets; Rome, Italy; 2020. 320 p. DOI: 10.4060/ca9692en

[84] Hirschi, K.D. Nutrient biofortification of food crops. Annu. Rev. Nutr. 2009;29:401-421. DOI: 10.1146/annurev-nutr-080508-141143

[85] Rengel, Z.; Batten, G.D.; Crowley, D.E. Agronomic approaches for improving the micronutrient density in edible portions of field crops. F. Crop. Res. 1999;60:27-40. DOI: 10.1016/S0378-4290(98)00131-2

[86] De Vita, P.; Platani, C.; Fragasso, M.; Ficco, D.B.M.; Colecchia, S.A.; Del Nobile, M.A.; Padalino, L.; Di Gennaro, S.; Petrozza, A. Selenium-enriched durum wheat improves the nutritional profile of pasta without altering its organoleptic properties. Food Chem. 2017;214:374-382. DOI: 10.1016/j.foodchem.2016.07.015

[87] Signore, A.; Renna, M.; D'Imperio, M.; Serio, F.; Santamaria, P. Preliminary evidences of biofortification with iodine of "Carota di Polignano", an Italian carrot landrace. Front. Plant Sci. 2018;9:170. DOI: 10.3389/fpls.2018.00170

[88] Nestel, P.; Bouis, H.E.; Meenakshi, J. V; Pfeiffer, W. Biofortification of staple food crops. J. Nutr. 2006;136:1064-1067. DOI: 10.1093/jn/136.4.1064.

[89] Bouis, H.E.; Saltzman, A. Improving nutrition through biofortification: A review of evidence from HarvestPlus, 2003 through 2016. Glob. Food Sec. 2017;12:49-58. DOI: 10.1016/j.gfs.2017.01.009

[90] Gregorio, G.B.; Senadhira, D.; Htut, H.; Graham, R.D. Breeding for trace minerals in rice. Food Nutr. Bull. 2000;21:382-386. DOI: 10.1177/156482650002100409

[91] Sanjeeva Rao, D.; Neeraja, C.N.; Madhu Babu, P.; Nirmala, B.; Suman, K.; Rao, L.V.S.; Surekha, K.; Raghu, P.; Longvah, T.; Surendra, P.; Kumar, R.; Babu, V.R.; Voleti, S.R. Zinc biofortified rice varieties: Challenges, possibilities, and progress in India. Front. Nutr. 2020;7:26. DOI: 10.3389/fnut.2020.00026

[92] Berardo, N.; Mazzinelli, G.; Valoti, P.; Laganà, P.; Redaelli, R. Characterization of maize germplasm for the chemical composition of the grain. J. Agric. Food Chem. 2009;57:2378-2384. DOI: 10.1021/jf803688t

[93] Ingallina, C.; Maccelli, A.; Spano, M.; Matteo, G. Di; Sotto, A. Di; Giusti, A.M.; Vinci, G.; Giacomo, S. Di; Rapa, M.; Ciano, S.; Fraschetti, C.; Filippi, A.; Simonetti, G.; Cordeiro, C.; Silva, M.S.; Crestoni, M.E.; Fornarini, S.; Mannina, L. Chemico-biological characterization of torpedino di fondi® tomato fruits: A comparison with san marzano cultivar at two ripeness stages. Antioxidants. 2020;9:1027. DOI: 10.3390/antiox9101027

[94] Hurtado, M.; Vilanova, S.; Plazas, M.; Gramazio, P.; Andújar, I.; Herraiz, F.J.; Castro, A.; Prohens, J. Enhancing conservation and use of local vegetable landraces: the Almagro eggplant (*Solanum melongena* L.) case study. Genet. Resour. Crop Evol. 2014;61:787-795. DOI: 10.1007/s10722-013-0073-2. The

[95] Lacchini, E.; Kiegle, E.; Castellani, M.; Adam, H.; Jouannic, S.; Gregis, V.; Kater, M.M. CRISPR-mediated accelerated domestication of African rice landraces. PLoS One. 2020;15:1-12. DOI: 10.1371/journal.pone.0229782

[96] Garg, M.; Sharma, N.; Sharma, S.; Kapoor, P.; Kumar, A.; Chunduri, V.; Arora, P. Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. Front. Nutr. 2018;5:12. DOI: 10.3389/fnut.2018.00012

[97] Schouten, H.J.; Krens, F.A.; Jacobsen, E. Cisgenic plants are similar to traditionally bred plants. EMBO Rep. 2006;7:750-753. DOI: 10.1038/sj.embor.7400769

[98] Panel, E. Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. EFSA J. 2012;10:10. DOI: 10.2903/j.efsa.2012.2943

[99] Holme, I.B.; Wendt, T.; Holm, P.B. Intragenesis and cisgenesis as alternatives to transgenic crop development. Plant Biotechnol. J. 2013;11:395-407. DOI: 10.1111/pbi.12055

ANNEX 4

Medina-Lozano, I, Bertolín, JR, Zufiaurre, R, Díaz, A (2020). Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.). *J. Vis. Exp.* 160, e61440. https://doi.org/10.3791/61440.

jove

Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.)

Inés Medina-Lozano[1,4], **Juan Ramón Bertolín**[2,4], **Raquel Zufiaurre**[3], **Aurora Díaz**[1,4]

[1] Unidad de Hortofruticultura, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA) [2] Unidad de Producción y Sanidad Animal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA) [3] Dpto Química Analítica, Escuela Politécnica Superior (Universidad de Zaragoza) [4] Instituto Agroalimentario de Aragón - IA2 (CITA-Universidad de Zaragoza)

[*] These authors contributed equally

Corresponding Author

Aurora Díaz
adiazb@cita-aragon.es

Citation

Medina-Lozano, I., Bertolín, J.R., Zufiaurre, R., Díaz, A. Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.). *J. Vis. Exp.* (160), e61440, doi:10.3791/61440 (2020).

Date Published

June 30, 2020

DOI

10.3791/61440

URL

jove.com/video/61440

Abstract

Vitamins, especially vitamin C, are important micronutrients found in fruits and vegetables. Vitamin C is also a major contributor to their antioxidant capacity. Lettuce is one of the most popular vegetables among consumers worldwide. An accurate protocol to measure vitamin C content in lettuce and other related species is crucial. We describe here a method using the ultra-high-performance liquid chromatography-ultraviolet (UPLC-UV) technique, in which sample preparation, vitamin extraction and chromatography conditions were optimized.

Samples were collected to represent the entire plant, frozen at -80 °C and lyophilized to prevent undesirable oxidation and make their manipulation easier. The extraction of vitamin C was carried out in acidic media, which also contributed to its stability. As vitamin C can be present in two different interconvertible forms, ascorbic acid (AA) and dehydroascorbic acid (DHAA), both compounds should be measured for accurate quantification. The DHAA was quantified indirectly after its reduction to AA because AA shows a higher absorptivity than DHAA in the UV range of the spectrum. From the same extract, two measurements were carried out, one before and one after that reduction reaction. In the first case, we were quantifying the AA content, and in the second one, we quantified the sum of AA and DHAA (TAA: total ascorbic acid) in the form of AA. Then, DHAA quantity was indirectly obtained by subtracting AA coming from the first measurement from TAA. They were determined by UPLC-UV, using a commercial AA standard to build a calibration curve and optimizing the chromatographic procedure, to obtain AA peaks that were completely resolved in a short time. This protocol could

be easily extrapolated to any other plant material with slight or no changes. Its accuracy revealed statistically significant differences otherwise unperceived. Other strengths and limitations are discussed more in depth in the manuscript.

Introduction

Cultivated lettuce (*Lactuca sativa* L.) is one of the most produced and consumed leafy vegetables worldwide, with a total production of about 27.3 million tons in 2018[1]. Lettuce is perceived as healthy by consumers. The nutritional properties are mainly attributed to the source of antioxidant compounds in the crop, such as vitamin C, among others like polyphenols and vitamin E[2]. Vitamin C is an essential micronutrient for humans unlike many other vertebrates, as we are unable to produce it due to mutations present in the gene coding for the last step enzyme in the biosynthetic pathway[3]. It is required for a normal cell metabolism and it also plays an important role in immune responses mainly due to its antioxidant activity[3,4].

Total vitamin C is made up of ascorbic acid (AA) and dehydroascorbic acid (DHAA). AA is the most biologically active form of the vitamin, but DHAA (its oxidation product) also shows biological activity and it can be easily converted into AA in the human body[5]. Therefore, quantifying both forms is important to determine the total vitamin C content of any horticultural crop, lettuce included.

A wide variety of approaches based on different analytical techniques have been used to measure vitamin C in vegetables, such as enzymatic, spectrophotometric, and titrimetric methods[6,7,8]. Although these methods are simple, they are not chemically specific for AA[9]. Consequently, chromatographic methods are preferred, especially the high-performance liquid chromatography-ultraviolet (HPLC-UV) technique, because of their higher accuracy[10]. HPLC-UV has been used to determine vitamin C in a great diversity of crops, like broccoli, spinach and lettuce[11,12,13]. However, the simultaneous quantification of AA and DHAA is complicated due to the low absorptivity of DHAA in the UV range of the spectrum. Alternatively, DHAA can be determined indirectly by using a reducing agent that converts DHAA to AA, measuring total ascorbic acid (TAA), and then calculating the difference between TAA and AA. Due to the necessity of a reduction reaction, in some studies, only AA has been quantified[14], which could actually represent an underestimation of vitamin C activity. That additional reduction reaction is also needed to determine DHAA indirectly even when the last advance in liquid chromatography techniques, ultra-high performance liquid chromatography (UPLC), is used. That step also benefits from the advantages that UPLC exhibits when compared to HPLC: higher efficiency and resolution, increased sensitivity, shorter time analysis and lower solvent consumption[15]. In consequence, UPLC-UV technique has been utilized to quantify vitamin C in different crops[16].

In addition, AA is a very labile molecule; thus, it is important to develop a protocol that prevents its degradation during lettuce storage and vitamin C analysis[9]. In this context, the following protocol offers a rapid and improved quantification of vitamin C content in lettuce by UPLC-UV, as well as an efficient extraction procedure. Not only elite cultivars have been included in the present study, but also traditional landraces and some wild relatives due to their potential interest in crop breeding, specifically in the improvement of the nutritional value of lettuce.

Protocol

1. Plant material preparation

1. Sample at least two leaves per plant in 50 mL polypropylene tubes, an outer (older) and an inner (younger) one in order to represent more accurately the whole plant. Collect at least three biological replicates for each sample.
2. Freeze them immediately using liquid nitrogen and store them at -80 °C until use. Make sure the liquid nitrogen does not get into the tubes; otherwise they could explode when removed due to the gas expansion during vaporization.
   CAUTION: Gloves and a face shield are required due to the potential hazards associated with using liquid nitrogen.
3. Remove the caps from the tubes and place them on the trays within the freeze dryer chamber of the lyophilizer (**Table of Materials**) programmed as follows: -25 °C for 72 h, -10 °C for 10 h, 0 °C for 10 h, and 20 °C for at least 4 h. Maintain the condenser temperature and the vacuum constant during the freeze-drying process at -80.2 °C and 112 mTorr, respectively.
4. When the material is completely dry (between 4 and 7 days depending on the plant and the degree of compaction into the tube), preserve at 4 °C, -20 °C or -80 °C for short (days to weeks), medium (months) or long (years) storage, respectively. The inclusion of bags containing silica gel beads in the sample-containing tubes is recommended.
5. Place the lyophilized samples into 20 mL polypropylene tubes together with 10 mm diameter stainless steel balls and grind them with a multitube vortexer using the intensity and time needed to obtain a fine dust.
   NOTE: During the entire process, protect the samples from exposure to direct light.

2. Reagent and solution preparation

1. Prepare the solvent extraction solution: 8% acetic acid (v/v), 1% MPA (meta-phosphoric acid) (w/v), 1 mM EDTA (ethylenediaminetetraacetic acid).
   1. Calculate the total volume of solvent needed to process the whole set of samples taking into account that 5 mL will be added to each. To prepare 1 L of the solution, add to a flask: 30 g of MPA, 0.372 g of EDTA dehydrate, 80 mL of acetic acid and 500 mL of ultrapure water (scale volumes and quantities accordingly). Seal the flask mouth with plastic film.
   2. Once dissolved with the help of a magnetic stirrer, use a volumetric flask to accurately measure 1 L, adding the necessary ultrapure water.
2. Prepare the reduction reaction buffer (0.5 M Tris (2-amino-2-(hydroxymethyl)-1,3-propanediol) pH 9.0) and reducing solution (40 mM DTT (1,4-Dithiothreitol) with 0.5 M Tris pH 9.0).
   1. Calculate the total volume of reducing solution needed to process the whole set of samples taking into account that 200 µL will be added to each of them. To prepare 100 mL of the buffer, add to a beaker: 6.055 g of Tris and 90 mL of ultrapure water (scale volumes and quantities accordingly). Seal the beaker mouth with plastic film.
   2. Once dissolved with the help of a magnetic stirrer, adjust the solution to pH 9.0 by adding 2 M HCl and

use a volumetric flask to accurately measure 100 mL, adding the necessary ultrapure water.

3. To prepare 100 mL of the reducing solution, add to a beaker: 0.629 g of DTT (purity: 98%) and 90 mL of the buffer (0.5 M Tris pH 9.0) previously prepared (2.2.1 to 2.2.2). Scale volumes and quantities accordingly. Seal the beaker mouth with plastic film.

4. Once dissolved with the help of a magnetic stirrer, use a volumetric flask to accurately measure 100 mL, adding the necessary volume of buffer 0.5 M Tris pH 9.0.

   NOTE: The reducing solution is very unstable. That is why a freshly made solution is strongly recommended.

3. Sulphuric acid (0.4 M H_2SO_4)

   1. Calculate the total volume of 0.4 M sulphuric acid needed to process the whole set of samples taking into account that 200 µL will be added to each. To prepare 100 mL of the solution, add to a beaker: 80 mL of ultrapure water and then 2.22 mL of H_2SO_4 (purity: 96%, density: 1.84 g mL^{-1}). Use a volumetric flask to accurately measure 100 mL, adding the necessary ultrapure water.

      CAUTION: Sulphuric acid is very corrosive, so it must be handled using protective equipment and under hood. In addition, the acid should be added to ultrapure water, and not water to acid, to reduce fumes and avoid accidents.

4. Hydrochloric acid (2 M HCl).

   1. To prepare 100 mL of 2 M hydrochloric acid, add to a beaker: 80 mL of ultrapure water and then 6.13 mL of HCl (purity: 37%, density: 1.19 g mL^{-1}). Seal the beaker mouth with plastic film. Use a volumetric flask to accurately measure 100 mL, adding the necessary ultrapure water. Scale volumes accordingly.

      CAUTION: Hydrochloric acid is very corrosive, so it has to be handled using protective equipment and under hood. In addition, the acid should be added to ultrapure water, and not water to acid, to reduce fumes and avoid accidents.

5. AA standard (stock and dilutions)

   1. Weigh exactly 10 mg of AA standard (purity: 99%) using a precision balance and add 90 mL of mobile phase (ultrapure water pH 2.0 with formic acid).

   2. Once dissolved with the help of a magnetic stirrer, use a volumetric flask to accurately measure 100 mL, adding the necessary volume of ultrapure water pH 2.0 with formic acid.

      NOTE: Protect this stock solution from the exposure to light.

   3. Prepare five dilutions from the stock of the AA standard to obtain a calibration curve following the instructions in **Table 1** and proceed with step 5.2.

Standard	[AA] (µg mL^{-1})	AA (100 µg mL^{-1}) solution (µL)	Mobile phase (µL)[a]
1	0.5	5	995
2	2.5	25	975
3	5	50	950

4	10	100	900
5	25	250	750
[a] Ultrapure water pH 2.0 acidified by formic acid.			

Table 1: Protocol to prepare five standards of AA (ascorbic acid). Volumes of solute and solvent to prepare each of the different concentrations of the standards are indicated.

3. Extraction of AA and DHAA

NOTE: It is recommended to work under conditions of low light intensity during the extraction steps.

1. To a 15 mL polypropylene centrifuge tube, add 50 mg of lyophilized ground sample and 5 mL of the extraction solvent (step 2.1).
2. Shake the mixture using a vortex for 5 s and then an orbital shaker for 10 min at 2000 rpm.
3. Introduce the tube in an ultrasonic bath for 10 min at room temperature with ultrasound activated.
4. Centrifuge at 4,000 x *g* for 10 min at 4 °C.
5. Take the supernatant, pass it through a 0.22 µm regenerated cellulose filter and store it in a 5 mL amber vial. This is Extract 1, which contains AA and DHAA.
   NOTE: The protocol can be paused here by freezing the extracts at -80 °C and protecting them from exposure to light as AA and DHAA are very unstable and degrade easily in the presence of light, at high temperatures or under oxidizing atmospheres (**Supplemental File 1**).

4. DHAA reduction to AA to extract TAA

1. Transfer 200 µL of Extract 1 to a 2 mL amber vial for liquid chromatography and add 200 µL of the reducing solution (step 2.2). Close the vial with a PTFE-silicone plug with pre-opening and shake it with a vortex for 5 s.
2. Allow the solution to stand for 30 min at room temperature and protect from light.
3. Add 200 µL of 0.4 M H_2SO_4 to stop the reaction and stabilize AA in acidic pH. The resulting solution is Extract 2, which contains only AA and is actually TAA.

5. Determination

1. UPLC-UV preparation
   1. Prepare the working solutions described in **Table 2**, suitably filtered through 0.22 µm filters, sonicated for at least 10 min and place them in the UPLC system.
   2. Switch on the three UPLC modules and wait for the internal calibration process to finish.
   3. Open the software (e.g., Empower 3) and load the instrumental program described in **Table 2**: **Empower 3 | Run Samples | Vitamin C method | UPLC_PDA | Use QuickStart**.

4. Once the software is loaded with the correct program, access the UPLC management console: **Quaternary Solvent Manager | Click right mouse button | Launch Console**.
5. Proceed to the preparation and stabilization of the UPLC instrument: **System | Control | Startup**.
    1. Purge all UPLC lines for at least 5 min: **Prime Solvents | QSM | Check A, B, C, D and Seal Wash | Duration of prime > 5 min**.
    2. Purge and clean the injector: **Prime Solvents | SM | Check Wash solvent (> 45 s) and Check Purge solvent (> 35 cycles)**.
    3. Equilibrate UPLC to method conditions: **Equilibrate to Method | QSM | Flow (0.3 mL min^{-1}) | Solvent A (2%) | Solvent B (0%) | Solvent C (98%) | Solvent D (0%)**; **Equilibrate to Method | SM | Sample (5 °C) | Column (30 °C)** and **Equilibrate to Method | Other | Check Lamp On | Press Start**.
    4. Wait for at least 1 h (even more time is recommended) for the equipment to stabilize. Stability can be verified checking the pressure in the column in the **Launch Console**: **System | Quaternary Solvent Manager | QSM System Pressure**. Ensure that there are no identifiable trends in pressure changes (either increases or decreases) and the delta value is less than 10 psi.
    6. In the **QuickStart** screen, fill the matrix with the names of the standards and samples to be analyzed.
2. AA determination in the standards
    1. Transfer 1 mL of each of the five AA standards previously prepared (step 2.5.3) to 2 mL amber vials for liquid chromatography. Close the vial with a PTFE-silicone plug with pre-opening and inject 5 µL in the UPLC instrument.
    2. Carry out the chromatography following the procedure described in **Table 2** starting from most diluted to most concentrated.
3. AA determination in the samples
    1. Pipette 200 µL of Extract 1 in a 2 mL amber vial for liquid chromatography and add 800 µL of ultrapure water. Close the vial with a PTFE-silicone plug with pre-opening and inject 5 µL in the UPLC instrument.
    2. Carry out the chromatography following the procedure described in **Table 2**.
4. TAA determination in the samples
    1. Add 400 µL of ultrapure water to Extract 2. Close the vial with a PTFE-silicone plug with pre-opening and inject 5 µL in the UPLC instrument.
    2. Carry out the chromatography following the procedure described in **Table 2**.

Components and paramethers	Description
Instrument	Acquity UPLC H-Class
Detector	PDA eλ Detector λabs for AA=245 nm
Software	Empower 3

Column	Acquity UPLC HSS T3 (150 mm x 2.1 mm x 1.8 µm)
Channel A	CH_3OH
Channel B/Wash	H_2O:CH_3OH (50:50 v:v)
Channel C	Ultrapure water pH 2.0 acidified by formic acid[a]
Channel D/Seal Wash	Ultrapure water:acetonitrile (90:10 v:v)
Mobile phase	0.3 mL min^{-1} of 2%A + 98%C (isocratic mode)
Column temperature	30 °C
Autosampler temperature	5 °C
Injection volume	5 µL
AA retention time	1.874 min
Running time	3 min
[a] Undetermined volume of formic acid used until pH adjustment	

Table 2: Chromatographic procedure optimized to determine AA (ascorbic acid) in extracts from lettuce and wild relatives. Description of the components, conditions and solutions employed.

6. Quantification of AA and DHAA

1. Statistical analysis

1. Determine the analytical parameters of the chromatographic method as described by Bertolín et al.[18] (**Table 3**).

   NOTE: The values of the parameters presented in **Table 3** will need to be defined under specific experimental conditions.

Analytical parameters of the method	Values
Linear range (µg mL^{-1})	0.5-25
Linear equation	y=53,143.03x
R^2	0.99998
Limit of detection (mg AA g^{-1} of dry matter)	0.013
Limit of quantification (mg AA g^{-1} of dry matter)	0.045

Repeatability (CV, %)[a]	1.75
Intermediate precision (CV, %)[a]	4.22
Recovery (Rec, %)[b]	95.6±2.4
[a] CV: coefficient of variation	
[b] The recovery assay was performed with 10 aliquots containing 50 mg of the same sample, 5 spiked with 2 mg of AA g^{-1} of dry matter, and 5 non-spiked. %Rec=([AA]spiked sample-[AA]sample)/([AA]spiked)x100.	

Table 3: Optimized analytical parameters for the detection and quantification of AA (ascorbic acid) and TAA (total ascorbic acid). The linear range, the equation and the coefficient of determination of the calibration (R^2) curve, as well as the limits of detection and quantification of AA (the same for TAA), and the repeatability, intermediate precision and recovery were obtained with a sample injection volume of 5 µL.

2. Calculate the AA and TAA concentration.
    1. Open the standard and sample chromatograms: **QuickStart | Browse Project | Channels | "name of standard or sample" | PDA Ch1 245 nm@1.2 nm**.
    2. Integrate the corresponding peak (AA or TAA) in the standards and samples by clicking on its starting point (approximately 1.790 min) and dragging it with the mouse to its end point (approximately 1.910 min).
    3. Build a calibration curve representing the absorbance values determined chromatographically (step 5.2.) against the concentration of the five AA standards prepared above (**Table 1**).
    4. Interpolate the absorbance values of the samples determined in steps 5.3 and 5.4 and obtain the AA and TAA concentration, respectively, with the following formula:

$$y=mx+n$$

where *y* is the integrated peak area, *x* is the AA or TAA concentration in ppm and *m* and *n* are the slope and the *y*-intercept of the obtained regression line, respectively.

    5. For calculating the concentration of DHAA, apply the following formula:

$$[\mathrm{DHAA}](\mu\mathrm{g\ mL^{-1}}) = [\mathrm{TAA}] - [\mathrm{AA}]$$

NOTE: To obtain the total concentrations of the DHAA, AA and TAA in mg g^{-1} of dry weight, the values obtained directly interpolating in the calibration curve will have to be multiplied by the total extract volume and the dilution factor applied, and then divided by the weight of the sample used to carry out the extraction.

Representative Results

Vitamin C quantification in *Lactuca* matrixes requires the development of a chromatographic approach that can ensure reliable results. **Figure 1A** shows a chromatogram resulting from a non-optimized protocol (**Supplemental File 2**), which presents an AA peak together with an unidentified minor "shoulder". Nevertheless, after improving the extraction and chromatographic conditions, a resolved AA peak without

interferences of unknown compounds was achieved (**Figure 1B**). In addition, the use of UPLC-UV equipment instead of HPLC-UV allowed us to reduce the retention time (RT) for AA: 1.874 min in the optimized chromatograms *versus* 2.980 min in the non-optimized ones (**Figure 1**), as well as the running times, 3 and 7 minutes for the optimized and non-optimized protocols, respectively.

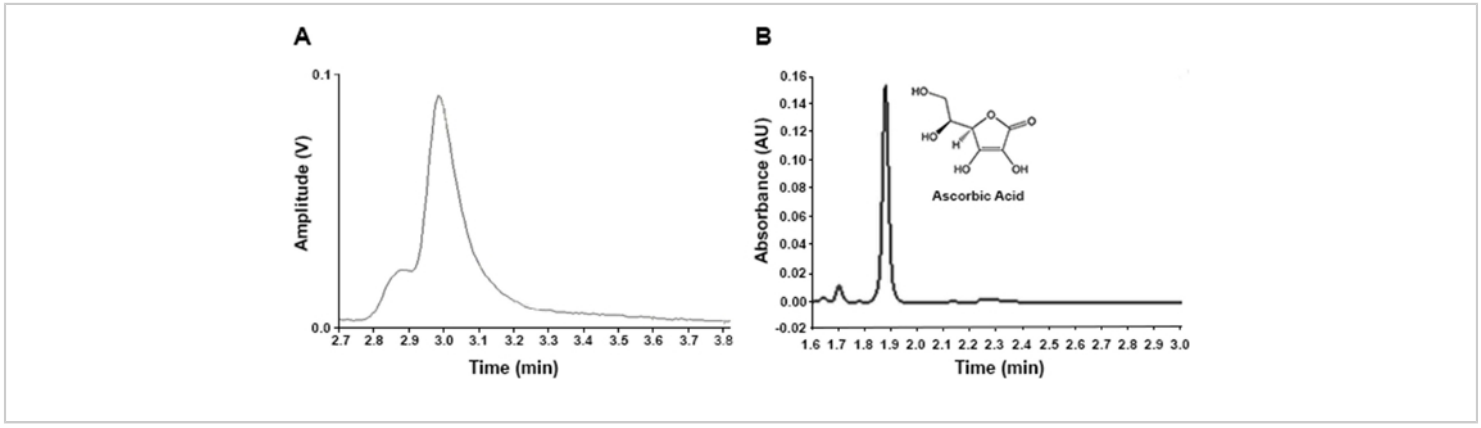


Figure 1: Chromatograms of AA in the same lettuce sample (commercial cultivar 'Begoña'). (**A**) HPLC-UV chromatogram resulting from a non-optimized protocol (conditions described in Supplemental File 2). (**B**) UPLC-UV chromatogram obtained with the optimized protocol (conditions described in **Table 2**). Please click here to view a larger version of this figure.

Interferences in AA peaks, like those observed in **Figure 1A**, consistently resulted in underestimation of vitamin C (AA, DHAA and TAA) content (**Figure 2**) due to an insufficient separation during the chromatographic process as the overlapping peak areas were integrated by a vertical drop at the deepest point between them. This bias is especially noticeable in the case of the crop wild relatives, particularly in DHAA and TAA content (**Figure 2**).

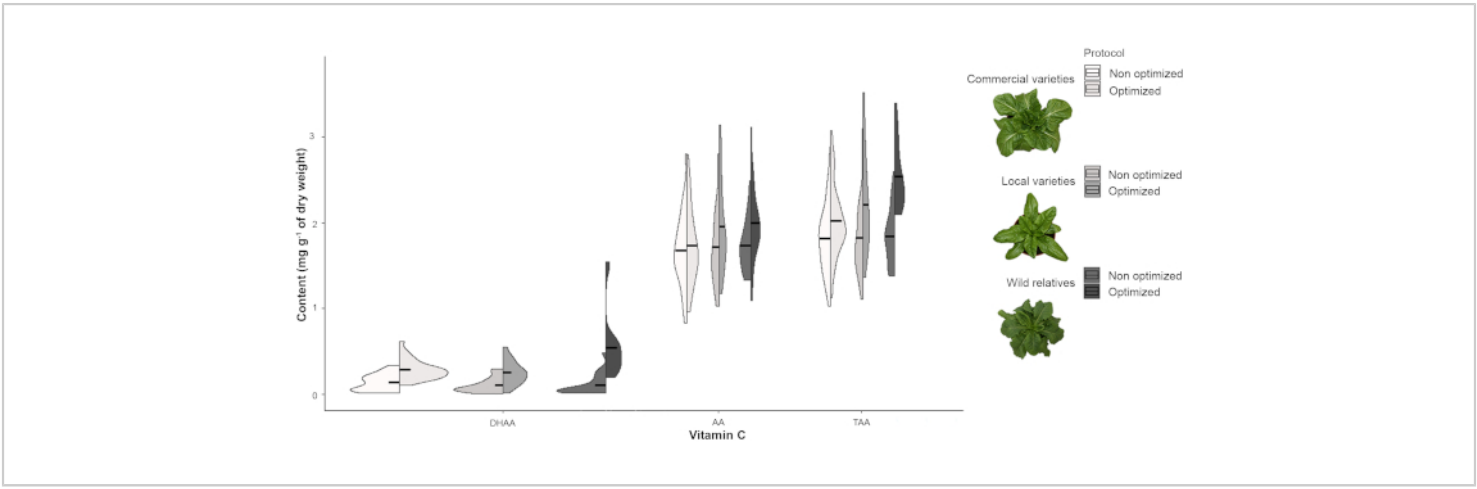


Figure 2: Distribution of the content of vitamin C. Split violin plots of DHAA, AA and TAA content (mg g^{-1} of dry weight) in commercial and traditional lettuce varieties and some wild relatives using non-optimized and optimized protocols. Black lines show the mean values. Please click here to view a larger version of this figure.

Furthermore, the use of a non-optimized protocol prevented us from extracting any useful conclusion from the results as they showed all samples, both types of lettuces and the wild relatives, having a similar vitamin C content. In contrast, the optimized protocol allowed us to detect statistically significant differences among them for DHAA and TAA content (**Table 4**), the richest ones being the wild species (**Figure 2**).

	Non optimized		**Optimized**	
	F-ratio	*p*-value	F-ratio	*p*-value
DHAA	0.460	0.637^{ns}	5.613	0.009**
AA	0.070	0.932^{ns}	1.020	0.374^{ns}
TAA	0.015	0.985^{ns}	4.438	0.022*
^{ns} , * and ** indicate non significant and significant at p<0.05 and 0.01, respectively.				

Table 4: Variation in the content of vitamin C. F-ratios (quotients of two variances, the between-group variance and the within-group variance) and significance values from the one-way ANOVA considering the type of *Lactuca* (commercial lettuce varieties, traditional lettuce varieties, and crop wild relatives) for DHAA, AA and TAA content in non-optimized and optimized protocols.

Supplemental File 1: AA and TAA stability at 5 °C over 24 h. (**A**) AA and TAA peak areas throughout 24 h. (**B**) AA and TAA content (mg g^{-1} of dry weight) throughout 24 h. Bars represent the standard deviations of two technical replicates

(n=2) kept in the autosampler at 5 °C and protected from exposure to light. Please click here to download this file.

Supplemental File 2: Main differences between the optimized and the non-optimized protocol for TAA, AA and DHAA extraction and quantification. The samples used were the same in both cases. Please click here to download this file.

Discussion

Vitamin C is a very important nutrient, but it is a very labile compound too, so its UPLC-UV quantification is dependent on multiple factors, such as sample storage and preparation, extraction method and chromatographic conditions. Therefore, a fast and simple procedure to prevent AA (with antioxidant power) oxidation to DHAA (without antioxidant properties) was needed. It was also crucial to avoid high pH and temperature conditions, as well as intense light and an oxidizing atmosphere during sample treatment to promote the stability of the compound.

To minimize AA oxidation, the following measures were taken. First of all, samples were lyophilized as a starting material for both protocols to ensure accurate quantification of vitamin C content and to easily manipulate samples. This option was preferred over fine grinding, commonly found throughout the literature[19], as the dust thaws very quickly so the water becomes available again. During the extraction procedure, a higher volume of a more acidic solution (8% acetic acid and 1% MPA) was used as extractant in the optimized protocol (**Supplemental File 2**), which also acted as a stabilizer by preventing AA degradation. This solution also contained EDTA as a chelating agent to increase stabilization[16], unlike the extractant in the non-optimized protocol (**Supplemental File 2**). Moreover, we tested if the extraction procedure could be enhanced by using two consecutive extractions with 2.5 mL of extractant instead of a single one with 5 mL and under a N_2 atmosphere instead of the standard atmospheric conditions. The best results were reached using only one extraction under an unmodified atmosphere, which simplified the protocol by making unnecessary additional steps (data not shown). Other minor changes were also introduced in the protocol in order to enhance the extraction (i.e., sonication), obtain a clearer extract (finer filtration) and reduce the protocol duration (**Supplemental File 2**). Regarding the chromatographic conditions, the validation of the method was carried out as reported before[18], guaranteeing good analytical parameters (**Table 3**). Besides, the use of ultrapure water with formic acid (pH 2.0) and methanol (98:2 v:v) with a 0.3 mL min^{-1} flow, instead of monopotassium phosphate 30 mM (pH 3.0) at 1 mL min^{-1} as the mobile phase (**Supplemental File 2**), resulted in an improved method. The most important advancement was likely using a UPLC system instead of an HPLC, which allowed us greater control of impacting conditions (like the temperature) and resulting in resolved AA peaks without interferences by unknown compounds, in a shorter time and consuming less volume of extract (**Supplemental File 2**).

Nevertheless, there are two main limitations of this method. The first one is that DHAA cannot be measured directly using an UV detector due to its low absorptivity in the UV range of the spectrum. It is important to quantify the DHAA content because it presents certain biological activity and is easily convertible to AA in the human body[5]. For that, an additional reaction to reduce DHAA to AA is needed, together with a second chromatographic run in order to measure TAA and then determine DHAA indirectly by subtracting AA content from TAA (**Figure 3**). In this sense, the reduction

step has been optimized by using a higher concentration of the reducing agent (DTT), increasing the reaction time from 5 to 30 min, and stopping the reaction with sulfuric acid (**Supplemental File 2**). The low stability of AA constitutes the second limitation of the method. As AA starts to degrade 4 h after extraction (**Supplemental File 1**), it is necessary to quantify it in this time interval. So, the number of samples to extract is conditioned by the chromatographic procedure. That is why we propose to freeze them at that step in this protocol, though in that case, not all of them could be placed in the UPLC autosampler to be measured automatically. Fortunately, the reduced RT for AA allowed us to obtain 3 min chromatograms, much shorter than the 7 min chromatograms obtained using HPLC (**Supplemental File 2**). Hence, vitamin C content could be determined in a high number of samples in a 4 h window.

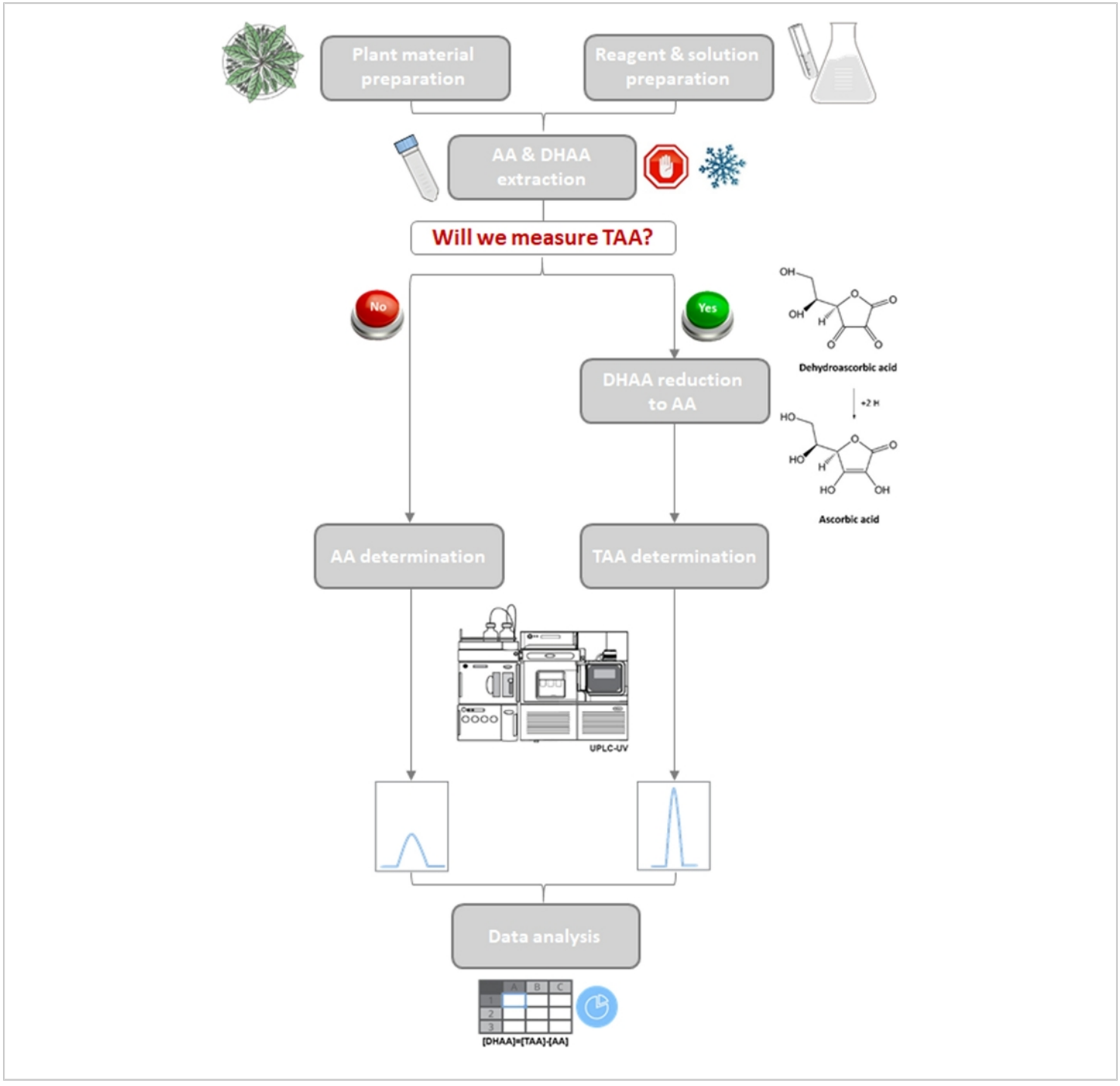


Figure 3: Workflow of the quantification of vitamin C in lettuce and some wild relatives.

Schematic diagram of the optimized protocol showing two branches for the determination of only AA or AA + DHAA (TAA).

Please click here to view a larger version of this figure.

As vitamin C is an essential nutrient for humans and due to its important health benefits, it has become the object of many studies. Therefore, it has been quantified in a great variety of crops, including lettuce, one of the most

consumed vegetables worldwide. Simple classical methods have been gradually replaced by liquid chromatography techniques because they are more specific and accurate[10]. However, due to the need of an additional reaction to quantify both, AA and DHAA via HPLC, in some studies on lettuce, only AA[14] or only TAA[11] (without quantifying AA before the reduction of DHAA into AA) have been measured. Furthermore, only a few authors have quantified AA and DHAA, despite the contribution of both molecules to vitamin C antioxidant activity[2]. Nevertheless, UPLC technique has become more important in recent times because of its higher performance when measuring vitamin C in several crops[16]. Comparing the results obtained in this study with the two methodologies, UPLC and HPLC, these advantages have been confirmed: well-defined AA peaks thanks to a higher sensitivity, and in very short times, have been achieved, which also implies fewer resources consumed. Despite of UPLC efficiency, only Chen et al.[20] have applied this technique to measure the vitamin C content in lettuce, which still led to an underestimation as only the AA form was quantified.

In summary, this work represents the first successful attempt to determine the total vitamin C content not only in different lettuce varieties but also in some of their wild relatives. Vitamin C quantification is also essential to select lettuces with higher antioxidant activity within breeding programs. In this sense, the increased total vitamin C content in lettuce wild relatives found here and the increased AA content reported in previous studies[14], as well as other antioxidant compounds[21], broadens the suitable candidates to improve the nutritional value of lettuces.

In conclusion, even with some limitations inherent to vitamin C's nature, like its gradual degradation few hours after being extracted or the need of a reduction reaction due to the low DHAA UV-absorptivity, it offers a less labor-intense and a less time-consuming method to measure vitamin C content. Additionally, it is also very robust and shows a high sensitivity and power of resolution. Moreover, it is easily transferable not only to other plant materials with slight or no changes, but also to processed products that supply the dietary intake of vitamin C to humans, which gives rise to a wide range of future applications in the emerging field of testing for reliable food quality.

Disclosures

The authors have nothing to disclose.

Acknowledgments

We thank J. A. Aranjuelo, A. Castellanos and "laboratorio de valoración nutritiva" from CITA for technical support and D. L. Goodchild for reviewing the English language. This work was funded by the projects RTA2017-00093-00-00 from the National Institute for Agricultural and Food Research and Technology (INIA) and LMP164_18 from the Government of Aragón; and by the Operational Programme FEDER Aragón 2014-2020 and the European Social Fund from the European Union [Grupos Consolidados A12-17R: "Grupo de investigación en fruticultura: caracterización, adaptación y mejora genéica" and A14-17R: "Sistemas agroganaderos alimentarios sostenibles" (SAGAS)]. I. M.L. was supported by a predoctoral contract for training doctors from the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Spanish State Research Agency (AEI).

References

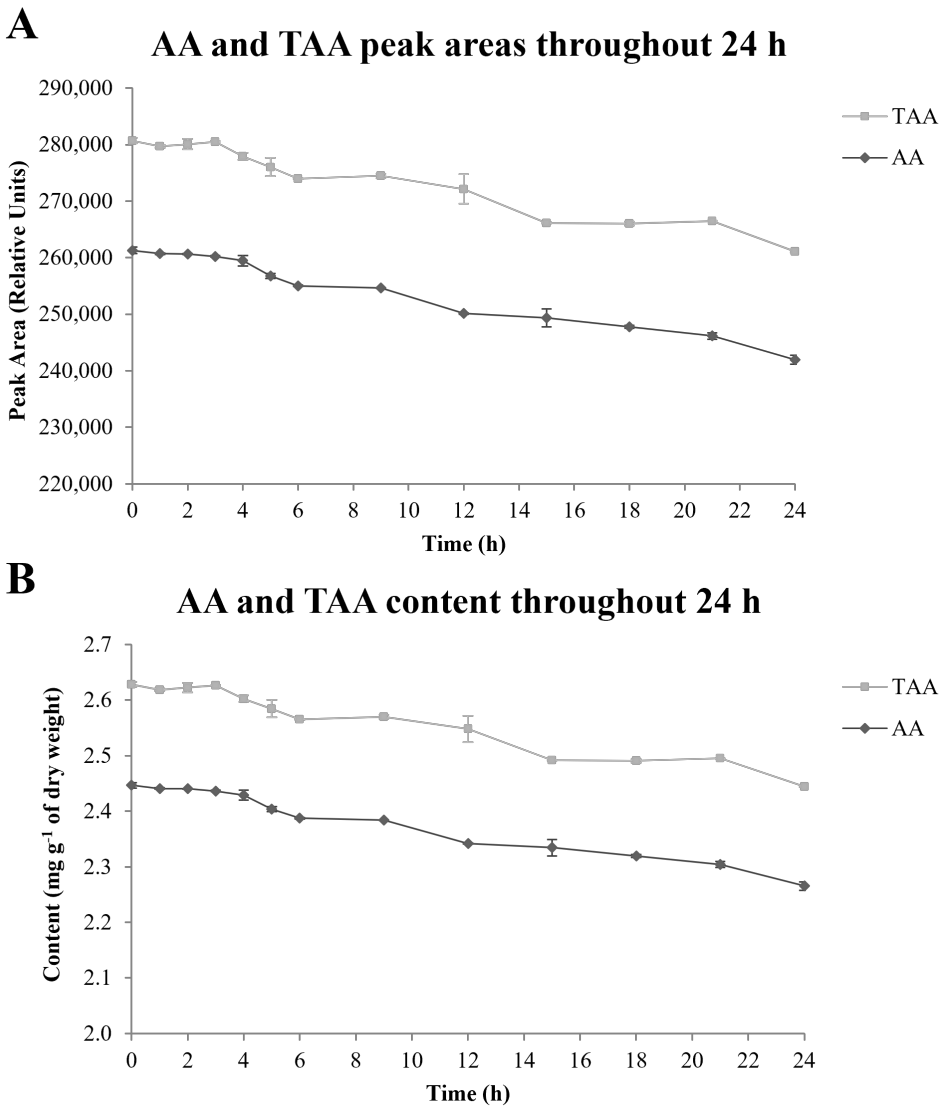
1. FAOSTAT. *Statistics of the Food and Agriculture Organization of the United Nations.* at <http://www.fao.org/faostat/en/#data/QC> (2018).
2. Llorach, R., Martínez-Sánchez, A., Tomás-Barberán, F. A., Gil, M. I., Ferreres, F. Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chemistry.* **108** (3), 1028–1038 (2008).
3. Carr, A.C., Frei, B. Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. *American Journal of Clinical Nutrition.* **69** (6), 1086–1107 (1999).
4. Carr, A.C., Vissers, M.C.M. Synthetic or food-derived vitamin C-Are they equally bioavailable? *Nutrients.* **5** (11), 4284–4304 (2013).
5. Lee, S.K., Kader, A.A. Preharvest and postharvest factors influencing vitamin C content of horticultural crops. *Postharvest Biology and Technology.* **20** (3), 207–220 (2000).
6. Shekhovtsova, T. N., Muginova, S. V., Luchinina, J. A., Galimova, A. Z. Enzymatic methods in food analysis: determination of ascorbic acid. *Analytica Chimica Acta.* **573–574**, 125–132 (2006).
7. Zhan, L. et al. Light exposure during storage preserving soluble sugar and L-ascorbic acid content of minimally processed romaine lettuce (*Lactuca sativa* L. var. longifolia). *Food Chemistry.* **136** (1), 273–278 (2013).
8. Malejane, D.N., Tinyani, P., Soundy, P., Sultanbawa, Y., Sivakumar, D. Deficit irrigation improves phenolic content and antioxidant activity in leafy lettuce varieties. *Food Science and Nutrition.* **6** (2), 334–341 (2017).
9. Tarrago-Trani, M.T., Phillips, K.M., Cotty, M. Matrix-specific method validation for quantitative analysis of vitamin C in diverse foods. *Journal of Food Composition and Analysis.* **26** (1–2), 12–25 (2012).
10. Klimczak, I., Gliszczyńska-Świgło, A. Comparison of UPLC and HPLC methods for determination of vitamin C. *Food Chemistry.* **175**, 100–105 (2015).
11. Złotek, U., Świeca, M., Jakubczyk, A. Effect of abiotic elicitation on main health-promoting compounds, antioxidant activity and commercial quality of butter lettuce (*Lactuca sativa* L.). *Food Chemistry.* **148**, 253–260 (2014).
12. Koh, E., Charoenprasert, S., Mitchell, A.E. Effect of organic and conventional cropping systems on ascorbic acid, vitamin C, flavonoids, nitrate, and oxalate in 27 varieties of spinach (*Spinacia oleracea* L.). *Journal of Agricultural and Food Chemistry.* **60** (12), 3144–3150 (2012).
13. Kałuzewicz, A. et al. The influence of short-term storage on the content of flavonoids and vitamin C in Broccoli. *European Journal of Horticultural Science.* **77** (3), 137–143 (2012).
14. van Treuren, R., van Eekelen, H.D.L.M., Wehrens, R., de Vos, R.C.H. Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. *Metabolomics.* **14** (11), 1–14 (2018).
15. Swartz, M.E. UPLC: An introduction and review. *Journal of Liquid Chromatography & Related Technologies.* **28** (7), 1253–1263 (2005).
16. Spínola, V., Mendes, B., Câmara, J.S., Castilho, P.C. An improved and fast UHPLC-PDA methodology for determination of L-ascorbic and dehydroascorbic acids

in fruits and vegetables. Evaluation of degradation rate during storage. *Analytical and Bioanalytical Chemistry.* **403** (4), 1049–1058 (2012).

17. Ball, G.F.M. *Water-soluble vitamin assays in human nutrition.* Springer, Boston, MA (1994).
18. Bertolín, J.R. et al. Simultaneous determination of carotenoids, tocopherols, retinol and cholesterol in ovine lyophilised samples of milk, meat, and liver and in unprocessed/raw samples of fat. *Food Chemistry.* **257**, 182–188 (2018).
19. Spínola, V., Llorent-Martínez, E. J., Castilho, P. C. Determination of vitamin C in foods: Current state of method validation. *Journal of Chromatography A.* **1369**, 2–17 (2014).
20. Chen, Y. et al. UVA radiation is beneficial for yield and quality of indoor cultivated lettuce. *Frontiers in Plant Science.* **10**, 1–10 (2019).
21. Damerum, A. et al. Elucidating the genetic basis of antioxidant status in lettuce (*Lactuca sativa*). *Horticulture Research.* **2** (15055), 1–13 (2015).

ANNEX 4. SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.jove.com/t/61440/improved-uplc-uv-method-for-quantification-vitamin-c-lettuce.



Supplemental File 1. AA and TAA stability at 5 °C over 24 h. (**A**) AA and TAA peak areas throughout 24 h. (**B**) AA and TAA content (mg g^{-1} of dry weight) throughout 24 h. Bars represent the standard deviations of two technical replicates (n=2) kept in the autosampler at 5 °C and protected from exposure to light.

Supplemental File 2. Main differences between the optimized and the non-optimized protocol for TAA, AA and DHAA extraction and quantification. The samples used were the same in both cases.

Section	Step	Optimized protocol	Non-optimized protocol
Reagent and solution preparation	Extraction solution	8% acetic acid (v/v), 1% MPA (w/v), 1 mM EDTA	5% acetic acid (v/v), 3% MPA (w/v)
	Reducing solution	40 mM DTT in 0.5 M Tris pH 9.0	30 mM DTT in dH_2O
	Sulphuric acid (H_2SO_4)	0.4 M	-
	Hydrochloric acid (HCl)	2M	5M
	Stock AA standard preparation	Solvent: Ultrapure water pH 2.0 acidified by 98-100% formic acid	Solvent: dH_2O
	AA calibration curve	5 points (0.5, 2.5, 5, 10, 25 µg mL^{-1})	6 points (10, 20, 40, 60, 80, 100 µg mL^{-1})
		Solvent: Ultrapure water pH 2.0 acidified by 98-100% formic acid	Solvent: 1.5% MPA
Extraction of AA and DHAA	Extraction solution volume (mL)	5	3
	Mixing	Vortex 5 s; orbital shaker (2000 rpm) 10 min at room temperature	Vortex 1 min
	Sonication (ultrasound bath)	10 min at room temperature	No
	Centrifugation (4,000 x *g*)	10 min at 4 °C	20 min at room temperature
	Filtration	0.22-µm regenerated cellulose filter	0.45-µm regenerated cellulose filter
DHAA reduction to AA	Incubation at room temperature	30 min	5 min
	Stop reaction	200 mL of 0.4 M H_2SO_4	No
Determination	Instrument	UPLC: Acquity H-Class	HPLC: Hewlett Packard 1050
	Detector	PDA eλ Detector λabs for AA=245 nm	eλ Detector λabs for AA=265 nm
	Column	Acquity UPLC HSS T3 (150 mm x 2.1 mm x 1.8 µm)	HPLC C18 Tracer column (250 mm x 4 mm x 5 µm)
	Channels	A: CH_3OH; B/Wash: H_2O:CH_3OH (50:50 v:v); C: Ultrapure water pH 2.0 with formic acid; D/Seal Wash	No
	Mobile phase	0.3 mL·min^{-1} of 2%A + 98%C (isocratic mode)	1 mL min^{-1} of KH_2PO_4 30 mM adjusted pH 3.0 HCl 5M (isocratic mode)
	Column temperature	30 ºC	Room temperature
	Autosampler temperature	5 ºC	No (room temperature)
	Injection volume (µL)	5	20
	AA retention time (min)	1.874	2.980
	Total running time (min)	3	7

ANNEX 5

Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2022). Validación de nuevos genes de referencia para estudios de expresión diferencial de genes involucrados en la síntesis de antocianinas en lechuga y especies silvestres relacionadas, in *Acta de Horticultura 90 (X Congreso Nacional de Mejora Genética de Plantas)*. Eds. R.A. Malvar, P. Fiz Rocha, (Pontevedra, Spain, Sociedad Española de Ciencias Hortícolas), 242-245.

57. Validación de nuevos genes de referencia para estudios de expresión diferencial de genes involucrados en la síntesis de antocianinas en lechuga y especies silvestres relacionadas

I. Medina-Lozano[1,2], M.S. Arnedo[3], J. Grimplet[1,2], A. Díaz[1,2,*]

[1]Departamento de Ciencia Vegetal, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Avd. Montañana 930, 50059, Zaragoza, España
[2]Instituto Agroalimentario de Aragón – IA2 (CITA-Universidad de Zaragoza), 50013, Zaragoza, España
[3]Ramiro Arnedo S.A. Paraje La Molina 54, Las Norias de Daza, 04716 Almería, España

Palabras clave: antioxidantes, estrés hídrico, *Lactuca sativa*, PCR cuantitativa a tiempo real, RNA-seq

RESUMEN

La lechuga (*Lactuca sativa* L.) es la hortaliza de hoja más popular a nivel mundial y su consumo sigue aumentando. Es además fuente de compuestos con efectos beneficiosos para la salud, como las antocianinas, potentes antioxidantes presentes en variedades de hoja roja o semirroja. En estudios de expresión génica mediante PCR cuantitativa (qPCR) es fundamental la selección de genes de referencia con expresión estable para la correcta normalización de los datos. En este trabajo se han seleccionado seis genes putativos de referencia para el análisis de la expresión diferencial de genes relacionados con la síntesis de antocianinas en tres estudios distintos: color de la hoja en variedades comerciales (verde vs. roja); tejidos en una especie silvestre relacionada (hoja vs. tallo); y condiciones de estrés hídrico en una variedad comercial, una variedad tradicional y una especie silvestre emparentada (control vs. DH3 y DH4). Con los datos de RNA-seq de dichas muestras se filtraron los genes con expresión estable y alta cobertura de secuencia y se realizó un ANOVA entre los grupos dentro de cada ensayo, seleccionándose seis sin diferencias significativas: *ADF2*, *CYB5*, *iPGAM*, *SCL13*, *TRXL3-3* y *VHA-H*. Su expresión se validó mediante qPCR, ofreciéndose rankings de estabilidad calculada con los algoritmos geNorm, NormFinder y BestKeeper, así como con el método Delta Ct. Englobando todos los resultados, *CYB5* y *TRXL3-3* fueron los genes más estables en los ensayos de expresión génica según el color de la hoja y el tejido (aunque en orden inverso), y *TRXL3-3,* seguido de *ADF2,* en los estudios de estrés hídrico. Los datos de RNA-seq son una nueva fuente para seleccionar genes de referencia, aunque la validación mediante qPCR sigue siendo aconsejable. Este ranking propuesto de nuevos genes de referencia podría resultar útil para futuros estudios de expresión génica en *Lactuca* spp. en las condiciones aquí descritas.

INTRODUCCIÓN

En las últimas décadas los programas de mejora de cultivos hortícolas se han centrado en el aumento de la producción y en la incorporación de resistencias a enfermedades, a veces en detrimento de caracteres relacionados con la calidad nutricional. En este sentido, resulta interesante abordar el enriquecimiento en compuestos beneficiosos para la salud humana de la lechuga, una de las hortalizas más consumidas a nivel mundial.

Este es el caso de las antocianinas, que son compuestos polifenólicos con potente actividad antioxidante, presentes exclusivamente en variedades de hoja roja y semirroja, y con una distribución desigual dependiendo del tejido. Aunque se desconoce cuál es su papel, se ha comprobado que en algunos cultivos como la vid, su contenido aumenta con el estrés hídrico (Ju et al., 2019). Sin embargo, hasta la fecha, esto no se ha estudiado en lechuga.

Por otro lado, la selección y validación de genes de referencia con expresión estable es esencial para realizar una correcta normalización de los datos en estudios de expresión génica diferencial mediante qPCR a tiempo real. Así, el objetivo de este estudio ha sido la validación de genes de referencia preseleccionados a partir de datos de RNA-seq para el análisis de la expresión diferencial de genes relacionados con la síntesis de antocianinas en lechuga en tres ensayos distintos: *i*) comparación según el color de la hoja en dos variedades comerciales (verde vs. roja); *ii*) comparación de tejidos en una especie silvestre relacionada (hoja vs. tallo); *iii*) condiciones de estrés hídrico en una variedad comercial, una variedad tradicional y una especie silvestre de *Lactuca* (control vs. DH3 y DH4).

MATERIALES Y MÉTODOS

En la comparación de lechugas de hoja verde y roja se usaron las variedades comerciales 'Begoña' y 'Romired', respectivamente; en el estudio de distintos tejidos (hoja y tallo) se empleó la especie silvestre de crecimiento arbustivo *Lactuca squarrosa* (Thunb.) Miq.; y en el ensayo de estrés hídrico se utilizaron la variedad comercial roja 'Romired', la variedad tradicional semirroja 'Morada de Belchite' y la especie silvestre *Lactuca homblei* de Wild. En este último, los tratamientos consistieron en tres regímenes de riego distintos aplicados las 3 semanas previas a la recolección: C (control o riego a demanda), DH3 (semana 1: 450 mL; semanas 2-3: 150 mL) y DH4 (semanas 1-3: 0 mL). En todos los casos se cultivaron tres plantas por accesión y, en el ensayo de estrés hídrico, también por tratamiento (3 réplicas biológicas), siguiendo un diseño de bloques aleatorizados. Tras aproximadamente 3 meses, se recogieron dos hojas por planta (exterior e interior), además de tallo en el caso de *L. squarro a*.

Posteriormente se realizó la extracción de ARN total de las muestras liofilizadas utilizando el kit NZY Total RNA Isolation (NZYtech Lda.-Genes and Enzymes), seguida de una digestión con ADNasa empleando el kit Turbo DNA-free™ (Invitrogen). Se construyeron un total de 39 librerías de ADN copia (ADNC) procedentes de los tres estudios: hoja verde vs. hoja roja (2 accesiones x 3 repeticiones), hoja vs. tallo (2 tejidos x 3 repeticiones) y C vs. DH3 y DH4 (3 accesiones x 3 repeticiones x 3 tratamientos). Estas se secuenciaron mediante el protocolo TruSeq Stranded mRNA de Illumina en el Centro Nacional de Análisis Genómicos (CNAG-CRG, Barcelona), utilizándose la herramienta Galaxy (Afgan et al., 2018) para analizar las secuencias obtenidas.

A partir de los datos de RNA-seq se filtraron los genes con expresión estable (|log2(fold change)|<1 y *p-valor* ajustado>0,05) y con una cobertura de secuencia amplia. Por último, se realizó un ANOVA y una comparación de medias entre los grupos dentro de cada ensayo y se seleccionaron seis genes sin diferencias significativas: *ADF2* (actin-depolymerizing factor 2), *CYB5* (cytochrome B5), *iPGAM* (probable 2,3-bisphosphoglycerate-independent phosphoglycerate mutase), *SCL13* (scarecrow-like protein 13), *TRXL3-3* (thioredoxin-like 3-3) y *VHA-H* (V-type proton ATPase subunit H).

Se aisló el ARN mensajero (ARNm) del ARN total con el kit Dynabeads® mRNA DIRECT™ (Invitrogen) y se sintetizó el ADN_C con el kit NZY M-MuLV First-Strand cDNA Synthesis, separate oligos (NZYTech). Las reacciones de qPCR se llevaron a cabo en un volumen final de 12 µL (2 µL de una dilución 1:40 de ADN_C, 0,40 mM de cada primer y 1x de la máster mix NZYSupreme qPCR Green (2x), ROX plus (NZYTech)), en el equipo StepOnePlus™ (Applied Biosystems). Se realizaron dos réplicas técnicas.

Se obtuvieron rankings de los valores de estabilidad calculados con los algoritmos geNorm (Vandesompele et al., 2002), NormFinder (Andersen et al., 2004), BestKeeper (Pfaffl et al., 2004) y con el método Delta Ct (Silver et al., 2006); todos ellos integrados en el software RefFinder (Xie et al., 2012). Para BestKeeper y el método Delta Ct se utilizaron los valores medios de los Cq (quantification cycles), mientras que para geNorm y NormFinder se utilizaron valores de Cq corregidos (CqE) con los datos de la eficiencia (E) según la fórmula CqE=Cq(log(E)/log(2)) (Garrido et al., 2020). En todos los casos se usaron las medias de las dos réplicas técnicas para cada muestra.

RESULTADOS Y DISCUSIÓN

La utilización de CqE con los algoritmos geNorm y NormFinder resulta conveniente, ya que la eficiencia de la qPCR varía e influye en gran medida en los resultados. Así, los rankings de estabilidad obtenidos son distintos según el análisis (Tabla 1), como está ampliamente descrito en la literatura. Por ello, se han utilizado los resultados de los cuatro métodos para la obtención de un ranking comprehensivo (Tabla 2) que permite determinar cuáles son los genes de referencia más estables en cada ensayo, con este material y bajo estas condiciones concretas. En los ensayos de expresión génica según el color de la hoja y el tejido, los genes más estables fueron *CYB5* y *TRXL3-3* (en orden inverso); y en el de estrés hídrico, *TRXL3-3* también fue el gen más estable, seguido de *ADF2* (Tabla 2).

En lechuga se han descrito genes nuevos de referencia para estudios de respuesta a estreses abióticos (sequía, salinidad, rayos UV-C y metales pesados) mediante qPCR (Borowski et al., 2014), si bien esta selección no está basada en datos de RNA-seq. Este trabajo ofrece un ranking de nuevos genes de referencia según sus valores de estabilidad en el material ensayado, seleccionados a partir de datos de RNA-seq y validados mediante qPCR. Algunos genes resultan adecuados para todos los ensayos (ej.: *TRXL3-3*); mientras que otros son claramente descartables (ej.: *VHA-H*) (Tablas 1 y 2). El uso de RNA-Seq y qPCR resulta una estrategia eficaz y fiable para la selección de genes de referencia distintos a los que se han venido utilizando hasta ahora y que no siempre resultan adecuados.

AGRADECIMIENTOS

Agradecemos al Banco de Germoplasma de Especies Hortícolas de Zaragoza (BGHZ-CITA, España) y al Centre for Genetic Resources (CNG, Wageningen, Países Bajos) por suministrarnos las semillas. Las fuentes de financiación de este trabajo son: RTA2017-00093-00-00 (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria), LMP164_18, LMP148_21 y Grupo Consolidado A12-17R (Gobierno de Aragón). IML disfruta un contrato predoctoral para la formación de doctores del Ministerio de Ciencia, Innovación y Universidades y la Agencia Estatal de Investigación.

REFERENCIAS

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., et al. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46: W537–W544.

Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. 2004. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64: 5245–5250.

Borowski, J.M., Galli, V., da Silva Messias, R., Perin, E.C., Buss, J.H., dos Anjos e Silva, S.D., and Rombaldi, C.V. 2014. Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. Planta 239: 1187–1200.

Garrido, J., Aguilar, M., and Prieto, P. 2020. Identification and validation of reference genes for RT-qPCR normalization in wheat meiosis. Sci. Rep. 10: 2726.

Ju, Y., Yang, B., He, S., Tu, T., Min, Z., Fang, Y., and Sun, X. 2019. Anthocyanin accumulation and biosynthesis are modulated by regulated deficit irrigation in Cabernet Sauvignon (Vitis Vinifera L.) grapes and wines. Plant Physiol. Biochem. 135: 469–479.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T. 2004. Determination of most stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26: 509-515.

Silver, N., Best, S., Jiang, J., and Thein, S.L. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol. Biol. 7:33.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3.

Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. 2012. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol. Biol. 80: 75–84.

TABLAS

Tabla 1. Valores de estabilidad de los genes de referencia candidatos según los algoritmos geNorm (gN), NormFinder (NF) y BestKeeper (BK) y el método Delta Ct (ΔCt) en tres ensayos distintos: hoja verde vs. hoja roja, hoja vs. tallo y estrés hídrico.

	Verde vs. roja				Hoja vs. tallo				Estrés hídrico			
Gen	**gN**	**NF**	**BK**	**ΔCt**	**gN**	**NF**	**BK**	**ΔCt**	**gN**	**NF**	**BK**	**ΔCt**
ADF2	0,33	1,09	0,57	0,83	0,77	1,18	0,89	1,10	1,54	0,77	1,23	4,04
CYB5	0,33	0,81	0,53	0,83	0,77	1,34	0,77	1,03	1,98	2,91	1,50	4,07
iPGAM	1,33	1,49	0,72	1,28	1,22	1,55	0,47	1,08	3,40	2,81	1,11	4,01
SCL13	1,90	2,88	0,53	0,92	2,41	4,73	0,92	1,54	4,41	6,62	0,92	3,89
TRXL3-3	0,54	0,71	0,40	1,10	1,01	0,30	0,44	0,93	1,54	0,77	0,85	3,90
VHA-H	1,11	0,77	1,12	1,41	1,07	0,30	1,23	1,38	6,82	11,27	10,35	14,08

Tabla 2. Rankings de estabilidad de los genes candidatos para tres ensayos distintos: hoja verde vs. hoja roja, hoja vs. tallo y estrés hídrico.

Ranking	Verde vs. roja	Hoja vs. tallo	Estrés hídrico
1	*CYB5*	*TRXL3-3*	*TRXL3-3*
2	*TRXL3-3*	*CYB5*	*ADF2*
3	*ADF2*	*ADF2*	*iPGAM*
4	*SCL13*	*iPGAM*	*SCL13*
5	*VHA-H*	*VHA-H*	*CYB5*
6	*iPGAM*	*SCL13*	*VHA-H*

ANNEX 6

Medina-Lozano, I (2021). Variedades tradicionales, apuesta segura. *Red de Intercambio de Conocimiento Agroalimentario*.

Opiniones y Experiencias
rica Red de Intercambio de Conocimiento Agroalimentario

Variedades tradicionales, apuesta segura / Inés Medina

Fecha: 15-Mar-2021

Inés Medina Lozano
Unidad de Hortofruticultura
Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA)
Instituto Agroalimentario de Aragón (IA2)
imedina@cita-aragon.es

Antes de poner de manifiesto la importancia de las variedades tradicionales de cultivos agrícolas, cabe destacar que se definen, por lo general, como aquellas que han mantenido los agricultores a lo largo del tiempo mediante selección activa y/o mejora de las mismas en un área concreta y que, por tanto, están adaptadas a las condiciones ambientales locales. Asimismo, las variedades tradicionales constituyen un punto intermedio en la escala de la evolución entre las especies silvestres relacionadas, que podemos considerar como los ancestros, y las variedades comerciales, que todos conocemos y vemos cada día en los supermercados, y que han llegado hasta ahí gracias a procesos de selección y mejora de ciertos caracteres. Un ejemplo concreto de estos tipos de material vegetal, en este caso de lechuga, se puede ver en la Figura 1.

Figura 1. *Ejemplos de especie silvestre, variedad tradicional y variedad comercial de lechuga. Fuente: Proyectos RTA2017-00093-00-00 y LMP164_18.*

Las variedades tradicionales se caracterizan por constituir poblaciones muy diversas que albergan una enorme variabilidad genética. Estas propiedades les han permitido adaptarse a distintas condiciones locales y afrontar ciertos cambios adversos. Pero no solo eso: dentro de esta gran diversidad, también es más probable encontrar individuos que sobresalgan por encima del resto en cuando a propiedades nutricionales se refiere. No obstante, la importancia del valor nutricional de los cultivos se dejó de lado en los años 60 con el surgimiento de la Revolución Verde, que tenía como principal objetivo el aumento del rendimiento mediante la obtención de cultivos más productivos y/o más resistentes a enfermedades y plagas. Esto ha provocado en muchas ocasiones una reducción en la cantidad de nutrientes de frutas y hortalizas, contribuyendo así a la malnutrición, problema de salud pública a nivel global. De hecho, según datos de la FAO, casi 690 millones de personas, prácticamente el 9% de la población mundial, tuvieron problemas de desnutrición en 2019, principalmente en los países en vías de desarrollo. Sin embargo, esta deficiencia de nutrientes en frutas y hortalizas que consumimos habitualmente, que es en parte responsable de la conocida como hambre oculta, también afecta a países desarrollados.

Por lo tanto, si estos individuos más ricos nutricionalmente están a nuestro alcance, podemos tener en nuestras manos una buena alternativa tanto para ayudar a combatir el hambre oculta como para satisfacer nuestras preocupaciones actuales por una dieta saludable. Son muchos los casos en los que variedades tradicionales de cultivos agrícolas básicos son más ricas en nutrientes esenciales y/o en compuestos beneficiosos para la salud, como los antioxidantes, tan de moda hoy en día: variedades de tomate de la sierra del Segura (Negro Yeste o Verdal) más ricas en β-caroteno (precursor de la vitamina A) o variedades de tomate de Murcia (CIDA-62) más ricas en carotenoides, especialmente licopeno; variedades de melón de distintas zonas de la región de Madrid (Tendral Negro o Tradicional de Villaconejos) con alto contenido en vitamina C; variedades de trigo de Canarias (Raspinegro canario o Arisnegro de Tenerife) más ricas en minerales como hierro, cobre, zinc y manganeso; variedades de berenjena de Valencia (IVIA25) con polifenoles abundantes y un largo etcétera. Además, estos son solo ejemplos de variedades tradicionales cultivadas en España. A nivel mundial las opciones que tenemos son mucho mayores.

Por otro lado, las variedades tradicionales constituyen un buen punto de partida en programas de mejora tanto clásica como biotecnológica para la obtención de cultivos biofortificados, es decir, aquellos que presentan *per se* un valor nutricional más alto. En ambos casos pueden aportar variabilidad genética, escasa en las variedades comerciales, así

como suponer un ahorro considerable de tiempo y recursos al tratarse de transferencia de genes entre variedades de la misma especie (cisgénesis), en vez de entre especies silvestres relacionadas y variedades comerciales (transgénesis). Además, se cree que la cisgénesis tendrá más aceptación que la transgénesis entre los consumidores. Otra de las ventajas que presentan las variedades tradicionales es que podrían suscitar cierto interés en estos tiempos de cocina vanguardista, debido a su aspecto menos común y a sus sabores, en algún caso olvidados. Ejemplos de esto son el plátano rojo, procedente de Ecuador y que actualmente se comercializa en Canarias, del que se dice que presenta un sabor dulce que recuerda ligeramente a la frambuesa o las zanahorias moradas procedentes de Italia, ambos con un alto contenido en antioxidantes (Figura 2). Por último, el mercado de proximidad podría verse beneficiado, especialmente en áreas rurales, con el regreso de variedades que trabajaban las generaciones pasadas.

Figura 2. *Plátano rojo y zanahoria morada. Fuente: https://pixabay.com/es/*

Ahora bien, a pesar de sus cualidades, el interés en las variedades tradicionales por parte de los agricultores ha ido disminuyendo a lo largo de los años debido a la modernización e intensificación de la agricultura, viéndose, por tanto, sometidas a un importante proceso de desaparición y a una erosión genética sustancial a favor de nuevas variedades comerciales. Por suerte, el relevante papel de los bancos de germoplasma, encargados de conservar y catalogar material vegetal de todo tipo, ha permitido preservar gran parte de la biodiversidad de las plantas. Es más, las variedades tradicionales constituyen el tipo de material vegetal más abundante en estos bancos, con una representación del 37% de un total aproximado de 2,5 millones de entradas en Genesys (portal que aporta datos de caracterización y evaluación de recursos genéticos de plantas procedentes de bancos de germoplasma de todo el mundo). Estas están seguidas de material de mejora e investigación, variedades avanzadas y mejoradas y especies silvestres, con representaciones del 27%, 19% y 17%, respectivamente. Esto demuestra cómo los bancos de germoplasma han sabido ver lo valiosas que son las variedades tradicionales. No nos quedemos los demás atrás.

ANNEX 7

Journal metrics, subject areas and contribution of the doctoral student to the indexed publications that constitute this thesis.

Journal metrics, subject areas and contribution of the doctoral student to the indexed publications that constitute this thesis

Publication 1 (Chapter 1)	
Medina-Lozano, I, Bertolín, JR, Díaz, A (2021). Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* 359, 129864. https://doi.org/10.1016/j.foodchem.2021.129864.	
Journal: Food Chemistry	
Impact Factor (JCR): 9.231	**Quartile**: Q1 (D1)
Subject Area: Food science and technology	**Rank:** 8/144
Contribution of the doctoral student: Analysis, experimentation, writing - review and editing.	

Publication 2 (Chapter 2)	
Medina-Lozano, I, Bertolín, JR, Plieske, J, Ganal, M, Gnad, H, Díaz, A (2024b). Studies of genetic diversity and genome-wide association for vitamin C content in lettuce (*Lactuca sativa* L.) using high-throughput SNP arrays. *Plant Genome* 17, e20518. https://doi.org/10.1002/tpg2.20518.	
Journal: The Plant Genome	
Impact Factor (JCR) in 2023: 3.9	**Quartile**: Q1
Subject Area: Plant sciences	**Rank**: 49/265
Contribution of the doctoral student: Data curation, formal analysis, investigation, methodology, visualization, writing - original draft, writing - review and editing.	

Publication 3 (Chapter 3)	
Medina-Lozano, I, Bertolín, JR, Díaz, A (2024a). Impact of drought stress on vitamin C and anthocyanin content in cultivated lettuces (*Lactuca sativa* L.) and wild relatives (*Lactuca* spp.). *Front. Plant Sci.* 15, 3389. https://doi.org/10.3389/fpls.2024.1369658.	
Journal: Frontiers in Plant Science	
Impact Factor (JCR) in 2023: 4.1	**Quartile**: Q1
Subject Area: Plant sciences	**Rank**: 44/265
Contribution of the doctoral student: Data curation, formal analysis, investigation, methodology, writing - review and editing.	

Publication 4 (Chapter 4.1)	
Medina-Lozano, I, Arnedo, MS, Grimplet, J, Díaz, A (2023). Selection of Novel Reference Genes by RNA-Seq and Their Evaluation for Normalising Real-Time qPCR Expression Data of Anthocyanin-Related Genes in Lettuce and Wild Relatives. *Int. J. Mol. Sci.* 24, 3052. https://doi.org/10.3390/ijms24033052.	
Journal: International Journal of Molecular Sciences	
Impact Factor (JCR): 4.9	**Quartile**: Q1
Subject Area: Biochemistry and molecular biology	**Rank:** 66/313
Contribution of the doctoral student: Data curation, formal analysis, investigation, methodology, resources, software, validation, visualization, writing - original draft preparation, writing - review and editing.	

Publication 5 (Chapter 4.2)	
Medina-Lozano, I, Grimplet, J, Díaz, A (2025). Harnessing the diversity of a lettuce wild relative to identify anthocyanin-related genes transcriptionally responsive to drought stress. *Front. Plant Sci.* 15, 1494339. https://doi.org/10.3389/fpls.2024.1494339.	
Journal: Frontiers in Plant Science	
Impact Factor (JCR) in 2023: 4.1	**Quartile**: Q1
Subject Area: Plant sciences	**Rank**: 44/265
Contribution of the doctoral student: Data curation, formal analysis, investigation, methodology, software, validation, visualization, writing - original draft, writing - review and editing.	

Publication 6 (Annex 2)	
Medina-Lozano, I, Díaz, A (2022). Applications of Genomic Tools in Plant Breeding: Crop Biofortification. *Int. J. Mol. Sci.* 23, 3086. https://doi.org/10.3390/ijms23063086.	
Journal: International Journal of Molecular Sciences	
Impact Factor (JCR): 5.6	**Quartile**: Q1
Subject Area: Biochemistry and molecular biology	**Rank:** 66/285
Contribution of the doctoral student: Writing - original draft, writing - review and editing.	

Publication 7 (Annex 3)
Medina-Lozano, I, Díaz, A (2021). Nutritional Value and Phytochemical Content of Crop Landraces and Traditional Varieties, in *Landraces – Traditional Variety and Natural Breed*, ed. A. Elkelish (IntechOpen), 95–116. ISBN 978-1-83968-718-1. https://doi.org/10.5772/intechopen.95514.
Book: Landraces – Traditional Variety and Natural Breed
Subject Area: Life sciences
Contribution of the doctoral student: Writing - original draft, writing - review and editing.

Publication 8 (Annex 4)	
Medina-Lozano, I, Bertolín, JR, Zufiaurre, R, Díaz, A (2020). Improved UPLC-UV Method for the Quantification of Vitamin C in Lettuce Varieties (*Lactuca sativa* L.) and Crop Wild Relatives (*Lactuca* spp.). *J. Vis. Exp.* 160, e61440. https://doi.org/10.3791/61440.	
Journal: Journal of Visualized Experiments	
Impact Factor (JCR): 1.355	**Quartile**: Q3
Subject Area: Multidisciplinary sciences	**Rank**: 49/72
Contribution of the doctoral student: Analysis, data curation, experimentation, formal analysis, methodology, visualization, writing - original draft, writing - review and editing.	

GLOSSARY OF ABBREVIATIONS

3MaT	malonyl coenzyme A: anthocyanin 3-*O*-glucoside-6''-*O*-malonyltransferase
4CL	4-coumarate CoA ligase
AA	ascorbic acid
ABA	abscisic acid
ABCB13	ABC (ATP (adenosine triphosphate)-binding cassette) transporter B family member 13
AB	advanced backcross
Act	actin
ADF2	actin-depolymerizing factor 2
AFLP	amplified fragment length polymorphism
AMOVA	analysis of molecular variance
ANOVA	analysis of variance
ANS	anthocyanidin synthase
ANTR4;2C	anion transporter 4;2C chloroplastic
APS2	ATP sulfurylase 2
APX	ascorbate peroxidase
As1	asparagine synthase 1
AT	acetyltransferase
BC	backcross
BGHZ	Vegetable Germplasm Bank of Zaragoza
bp	base pair
C	control
C4H	cinnamate 4-hydroxylase
CAPS	cleaved amplified polymorphic sequence
cDNA	complementary deoxyribonucleic acid
CE	capillary electrophoresis
CGN	Centre for Genetic Resources
CHI	chalcone isomerase
CHS	chalcone synthase
CITA	Centro de Investigación y Tecnología Agroalimentaria de Aragón
CPM	counts per million
CPR1	constitutive expressor of PR (pathogenesis-related) genes 1
Cq	quantification cycle
CqE	Cq data corrected with the efficiency values
CR	call rate
Ct	threshold cycle
CV	coefficient of variation
CWR	crop wild relatives
CYB5	cytochrome B5
DArT-seq	diversity array technology sequencing
DEG	differentially expressed gene
DET1	de-etiolated 1
DFR	dihydroflavonol 4-reductase
DHAA	dehydroascorbic acid
DHAR	dehydroascorbate reductase
DHL	doubled haploid line
DI	deficit irrigation
DPH	diphthine methyl ester synthase
DTT	1,4-dithiothreitol
DW	dry weight

E1, E2, E3	extract 1, 2, 3
E2FC	elongation factor 2FC
EBG	early biosynthesis gene
EDTA	ethylenediaminetetraacetic acid
eEF1-α	eukaryotic elongation factor 1-α
EF2	elongation factor 2
EGL2	endoglucanase 2
ERF1	ethylene responsive factor 1
ESCRT	endosomal sorting complex required for transport
EXL3	extracellular lipase 3
EXT1	extensin-1-like
F3H	flavanone 3-hydroxylase
F3'H	flavanone 3'-hydroxylase
FAF1	fas-associated factor 1-like
FarmCPU	Fixed and random model Circulating Probability Unification
FBP	F-box protein
FC	fold change
FDR	false discovery rate
FGS	first-generation sequencing
FLD	Fisher's linear discriminant
FREE1	FYVE (Fab1 (1-phosphatidylinositol 3-phosphate 5-kinase) YOTB, Vac1 (vacuolar transport protein), and EEA1 (early endosome antigen 1)-domain protein required for endosomal sorting) 1
FRET	fluorescence resonance energy transfer
FST	Wright's fixation index
FST	flavonol 3-sulfotransferase
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GBS	genotyping-by-sequencing
GBSS	granule-bound starch synthase
GC	gas chromatography
GD	gene diversity
GEBV	genomic estimated breeding value
GDSL	glycine-asparagine-serine-leucine
GELP	GDSL esterase/lipase
Gly	glycosyl
GMO	genetically modified organism
GO	gene ontology
GP	genomic prediction
GP1	primary gene pool
GP2	secondary gene pool
GP3	tertiary gene pool
GS	genomic selection
GWAS	genome-wide association study
HAD	haloacid dehalogenase
Ho	observed heterozygosity
HomFLD	homozygous Fisher's linear discriminant
HPLC	high-performance liquid chromatography
HPPD	4-hydroxyphenylpyruvate dioxygenase
HRM	high-resolution melting
HSP	heat shock protein
HVA22G	HVA22 (*Hordeum vulgare* ABA-induced protein) 22-like protein G
IBS	identity by state

Indel	insertion-deletion
iPGAM	2,3-bisphosphoglycerate-independent phosphoglycerate mutase
ISSR	inter-simple sequence repeat
KASP	kompetitive allele-specific PCR (polymerase chain reaction)
KFB	F-box/kelch-repeat
LBG	late biosynthesis gene
LC	liquid chromatography
LCMS	LC/mass spectrometry
LD	linkage disequilibrium
LettuceGDB	Lettuce Genome Database
lncRNA	long noncoding ribonucleic acid
LOD	limit of detection
LOG7	cytokinin riboside 5'-monophosphate phosphoribohydrolase
LOQ	limit of quantification
LTPG20	lipid transfer protein GPI (glycosyl-phosphatidylinositol)-anchored 20
LYCb	lycopene b-cyclase
MAF	minor allele frequency
MAGIC	multi-parent advanced generation inter-cross
MAS	marker-assisted selection
MDHAR	monodehydroascorbate reductase
MDS	multidimensional scaling
mGWAS	metabolic/metabolomic GWAS
MIK2	MDIS (male discoverer) 1-interacting receptor like kinase 2
MNP	multiple nucleotide polymorphism
MPA	metaphosphoric acid
mQTL	metabolic QTL
MSG	multiplex shotgun genotyping
MSR	methionine sulfoxide reductases
MVB	multivesicular body
MYB	myeloblastosis
NAC	NAM (no apical meristem), ATAF (*Arabidopsis thaliana* activating factor) and CUC (cup-shaped cotyledon)
NADPHP450R1	NADPH (nicotinamide adenine dinucleotide phosphate)-cytochrome P450 reductase 1
NAM	nested association mapping
NGS	next generation sequencing
NIL	near isogenic line
NMR	nuclear magnetic resonance
Nos (gene)	nopaline synthase
OMT	*O*-methyltransferase
PAL	phenylalanine ammonia lyase
PCA	principal component analysis
PLIP	phospholipid-inositol phosphatase
PPE	pectinesterase/pectinesterase inhibitor
PTFE	polytetrafluoroethylene
Q	subpopulation membership coefficient
QC	quality control
qPCR	quantitative PCR
QQ	quantile-quantile
QTL	quantitative trait locus
r^2	correlation coefficient
R^2	coefficient of determination
RAD-seq	restriction site-associated DNA sequencing

RAPD	random amplified polymorphic DNA
Rec	recovery
REST-seq	restriction fragment sequencing
RFLP	restriction fragment length polymorphism
RG	reference gene
RIL	recombinant inbred line
RNA-seq	RNA sequencing
ROS	reactive oxygen species
RRL	reduced representation library
RT	retention time
RTL3	ribonuclease III-like protein 3
SA	salicylic acid
SAUR50	small auxin up-regulated RNA 50
SBE	single-base extension
SBG	sequence-based genotyping
SBT3	subtilisin-like protease 3
SCL13	scarecrow-like protein 13
SCOOP	serine-rich endogenous peptide
SD	standard deviation
SGS	second-generation sequencing
siRNA	small interfering RNA
SL	strigolactone
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
SPP	single position polymorphism
spp.	species
SSR	simple sequence repeat
SV	stability value
TAA	total ascorbic acid
TF	transcription factor
tGBS	tunable genotyping-by-sequencing
TGS	third-generation sequencing
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TRXL3-3	thioredoxin-like protein 3-3
Tub	tubulin
Ub	ubiquitin
UFGT	anthocyanidin 3-*O*-glucosyltransferase
UPLC	ultra performance liquid chromatography
URT1	UTP (uridine triphosphate):RNA uridylyltransferase 1
USDA	United States Department of Agriculture
UV	ultraviolet
VHA-H	V-type proton ATPase subunit H
VIT2	vacuolar iron transporter 2
WGCNA	weighted gene co-expression network analysis
WIT	WPP domain-interacting tail-anchored protein
WPP	tryptophan-proline-proline
XRN3	5'-3' exoribonuclease 3

cita
CENTRO DE INVESTIGACIÓN Y TECNOLOGÍA AGROALIMENTARIA DE ARAGÓN

ia2
Instituto Universitario de Investigación Mixto
Agroalimentario de Aragón

Universidad Zaragoza
1474

GOBIERNO DE ARAGON
Departamento de Empleo, Ciencia y Universidades

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
AGENCIA ESTATAL DE INVESTIGACIÓN