Received: 4 December 2024 | Revised: 27 February 2025 | Accepted: 6 March 2025

DOI: 10.1002/ppp3.70017

OPINION

Plants People Planet PPP Open Access

The potential of seedbank digital information in plant conservation

Roberta Gargiulo[1] | Sean Hoban[2,3] | Anne M. Visscher[1] | Pablo Gómez Barreiro[1] | Roberta L. C. Dayrell[1] | Kiran L. Dhanjal-Adams[1] | Alice R. Hudson[1] | Charlotte E. Seal[1] | Simon Kallow[1] | Jan Sala[1] | Lee Oliver[1] | Naomi Carvey[1] | Ted Chapman[1] | Efisio Mattana[1] | Clare Callow[1] | Christopher Cockel[1] | Michael F. Fay[1,4] | Juan Viruel[1] | Inna Birchenko[1] | Stephanie Miles[1] | Timothy Pearce[1] | Paul Kersey[1] | Elinor Breman[1] | Kate A. Hardwick[1]

[1]Royal Botanic Gardens, Kew, Richmond, TW9 3DS, UK

[2]Center for Tree Science, The Morton Arboretum, Lisle, Illinois, USA

[3]Committee on Evolutionary Biology, The University of Chicago, Illinois, USA

[4]School of Plant Biology, University of Western Australia, Crawley, Australia

Correspondence
Roberta Gargiulo, Royal Botanic Gardens, Kew, Richmond, TW9 3DS, UK.
Email: r.gargiulo@kew.org

Societal Impact Statement

Seedbanks are vital for biodiversity conservation, but their potential remains underutilised due to a limited understanding of the intraspecific genetic diversity they hold. By leveraging digitised data associated with seedbank collections, such as sampling locations, number of maternal plants and seed traits, we can attempt the estimation of genetic variation and identify gaps in collections, enabling better prioritisation of species for conservation efforts. These advancements can inform policy targets like those of the Kunming-Montreal Global Biodiversity Framework, promoting more effective conservation strategies. Digitisation and emerging machine-learning technologies offer scalable, cost-efficient solutions to enhance conservation knowledge, ensuring biodiversity resilience for future generations.

Summary

Seedbank collections hold significant untapped potential for advancing conservation science and practice, but the intraspecific genetic diversity (i.e. diversity within a species) stored in worldwide seedbank collections remains largely unknown, hindering the effective use of seeds for both informing and implementing in situ interventions. As producing genetic data is time-consuming and expensive, other data associated with seedbank collections can greatly enhance our understanding of the genetic variation stored in seed collections when genetic data are unavailable. Information such as the location of sampling sites, estimated population size and the number of mother plants from which seeds were collected can facilitate the estimation of the genetic diversity captured in the collections. This information can also be used to estimate the sampling effort required to fill gaps in seedbank collections to better represent genetic diversity,

---

Elinor Breman and Kate A. Hardwick are joint senior authors.

through comparison with existing baselines from species where genetic diversity is characterised, and through simulations. Digitisation of the data associated with seedbank collections makes the approaches above practicable at scale. In addition, digital images of the seeds themselves may identify intraspecific phenotypic variation and can, therefore, be used to prioritise populations for future genetic studies.

In this article, we explore the potential of digitised information made available by seedbanks for improving our understanding of the intraspecific genetic diversity preserved in collections. We describe possible improvements that might enhance the predictive power of digital information for genetic studies, and discuss the challenges and opportunities associated with these.

KEYWORDS
conservation genetics, digital data, high-resolution images, intraspecific genetic diversity, seedbank collections

1 | SEEDBANKS AS REPOSITORIES OF INTRASPECIFIC GENETIC DIVERSITY

Preserving germplasm ex situ is instrumental in safeguarding plant species against biodiversity loss. Among germplasm repositories, traditional seedbanks have proven to be a cost-effective and flexible solution for conserving most plant species (Li & Pritchard, 2009). As repositories of genetic diversity, these collections provide insurance against the loss of plant populations and the ecosystems to which they belong, represent a source of germplasm for future restoration efforts (Goodale et al., 2023), support human sustenance and well-being by preserving useful plant traits and provide living material to enable the study of these traits. In addition, and in the same fashion as museum and herbarium specimens (Burbano & Gutaker, 2023), seed collections made from populations at multiple time points can also provide information on intraspecific variability over time, and serve as baseline data to be compared with contemporary samples, especially in plants with short generation times. This can pave the way for characterising the evolutionary trajectories of populations (e.g., Franks et al., 2018) and for understanding the loss of genetic diversity over time for conservation (Gargiulo et al., 2024).

Conservation policy is progressively incorporating intraspecific genetic diversity within biodiversity monitoring frameworks (as for instance in the Kunming-Montreal Global Biodiversity Framework Target 4; Hoban et al., 2023; see also the Global Strategy for Plant Conservation Action 2, 3, 4c, 4d, 8b and 11, SBSTTA25/INF/4) for wild and cultivated or domesticated populations, and acknowledging its importance in allowing species survival in the face of environmental changes. However, collections of wild plant germplasm have historically been established following guidelines based on crops or plants with useful traits that are very well known (Hoban & Strand, 2015; Namoff et al., 2010), or based on general, rule-of-thumb sampling recommendations for capturing genetic diversity (Brown & Hardner, 2000; Brown & Marshall, 1995). For wild plant species, the actual intraspecific genetic diversity of collections has rarely empirically been demonstrated or compared to the overall genetic diversity of the species, and how representative collections remain largely unknown (but see Gargiulo et al., 2019; Hoban et al., 2020; Richards et al., 2010; Volk et al., 2005).

Characterising intraspecific genetic diversity in seedbank collections of wild plants would contribute to addressing several challenges in using seeds for research and practice, such as: (i) the identification of target populations to increase the representativeness of the collection, in terms of genetic diversity (Gargiulo et al., 2019; McGlaughlin et al., 2015; Namoff et al., 2010), including their adaptive genetic variation; (ii) the prediction of the outcome of restoration actions in terms of fitness of the restored populations and their capacity to adapt to a changing environment; and (iii) the selection of the material to ensure the most appropriate genetic composition in the restored populations (Bucharova et al., 2019; Borrell et al., 2020; Di Sacco et al., 2021; White et al., 2025).

In the case of crop seedbanks, which often include large collections of landraces, great progress has been made in understanding the genetic diversity they contain, with large-scale sequencing projects including rice, wheat, barley and most major crop species (Cheng et al., 2024; Milner et al., 2019; Sun et al., 2020; Wang et al., 2018), and interest in developing integrated information systems providing access to these data (Halewood et al., 2018). Often these projects have utilised SNP arrays or methods for reduced representation sequencing, customised for the species being studied (Edet et al., 2018). In the case of wild species, however, the number of seedbank accessions per species is typically much lower, while the number of species is greater; thus standardised (and cost-effective) laboratory and sequencing protocols are harder to achieve. Moreover, the full economic value of seeds of wild species is grossly underestimated and extraordinarily difficult to evaluate, beyond the main crops (Li & Pritchard, 2009; Mattana et al., 2022). Thus, the comprehensive sequencing of wild seedbank collections is not yet economically or practically feasible. How, therefore, can we understand whether the collections held are representative of intraspecific genetic diversity in

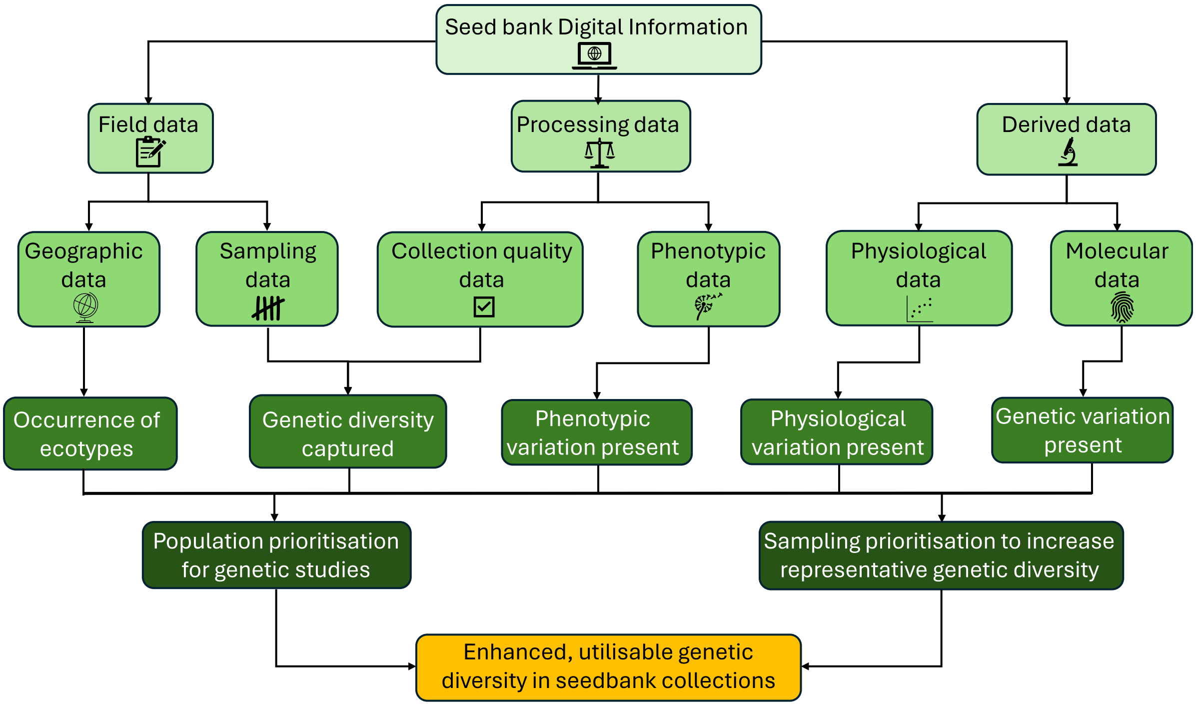

FIGURE 1 Categories of digital data associated with seedbank collections covered in this study and their potential value in informing conservation genetic science and practice.

the wild and suitable for restoration purposes? A potential approach is to select a subset of priority species for genetic study and complement these with the use of collection-associated data (available for many more species) to predict whether the intraspecific genetic diversity of other species has been captured and to prioritise further populations for genetic studies (Hoban, Volk, et al., 2018). In this perspective article, we will show how this can be done.

2 | THE VALUE OF DIGITAL DATA ASSOCIATED WITH SEEDBANK COLLECTIONS

There are three broad categories of digital data associated with seed collections (Figure 1):

(i) Field data: e.g. date, location, number of maternal plants, habitat descriptions.
(ii) Processing and associated data: e.g. photos of seeds, X-ray images (see Box 1), herbarium vouchers, seed numbers, collection quality and verified scientific name.
(iii) Derived data: e.g. data derived from analytical investigations and high-resolution images (e.g., morphological data, physiological data such as those associated with dormancy and germination requirements).

We will explore how some of these digital data can help the prediction of intraspecific genetic diversity (Figure 1).

2.1 | Number of mother plants, number of seeds and estimation of population census size

The estimated number of viable seeds, the number of maternal lines collected and the census size estimate of the population (i.e. number of reproductively mature individuals or N_C), when recorded, can be used as a very rough indicator of the genetic diversity within the collection, because on average, all else being equal, higher numbers of individuals will accumulate more genetic diversity. The sampling

Box 1 - Machine learning and seed image digitisation at the Millennium Seed Bank

The investment in acquiring data decreases dramatically when the focus transitions from major crops to less commercially important and wild species. In general, wild species, represent a less developed area of research, creating a unique niche in which evolving machine learning algorithms can not only automate data collection tasks but also facilitate the identification of easily measurable seed traits that serve as proxies for traits that are costly and time-consuming to assess. At the Millennium Seed Bank, several ongoing projects focus on the potential use of machine learning in combination with digital images to enhance research and conservation efforts (Figure 2).

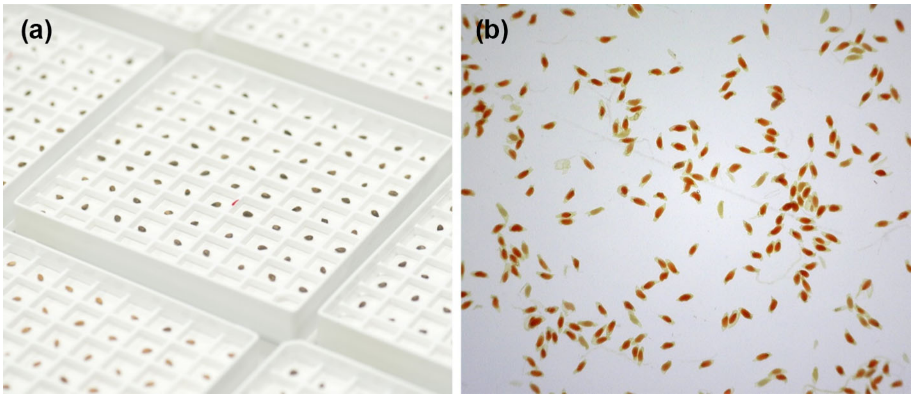

FIGURE 2 (a) *Pinus sylvestris* seeds (collected in the UK) and, (b) *Polystachya concreta* seeds (collected in Madagascar) stained in tetrazolium chloride, prior to digitisation for ongoing projects using machine learning to explore morphological seed characteristics and their relationship to germination and viability.

Case study 1: Exploring machine learning as a tool to improve the accuracy and speed of orchid viability testing

Orchid seeds are dust-like in size, and their reliance on specific symbiotic fungi for germination increases the effort required to assess the current quality of orchid collections. As a result, seedbanks holding orchid collections may not accurately record the true amount of viable material they hold, which could directly affect effort put into future collection priorities and hinder restoration outcomes for these species. The tetrazolium test is widely used among crop species to predict seed collection quality. In this test, seeds are imbibed in a solution of tetrazolium chloride, which turns red when it reacts with enzymes linked to respiration. Determining seed viability still relies on individuals interpreting the staining results, which can lead to inconsistencies between observers (e.g., Gosling, 2003; Pradham et al., 2022). The aim of this project is to reduce both the workload and subjectivity of the scoring process by using machine learning algorithms in combination with images of stained orchid seeds. The main goal is to develop a model that can automatically classify each image as viable, non-viable or empty, and make it accessible globally through an online application. This platform will also include a step-by-step protocol and an imaging kit to facilitate replication.

Case study 2: Using seed X-ray images to predict viability and potential for trait extraction.

X-ray images are routinely taken before seeds are banked. X-ray images allow the counting of the number of full, partially-full, empty or infested seeds in a sub-sample of the collection (usually 20–50 seeds) to estimate the proportion of potentially viable seeds in the overall collection. Although the process of totalling seed numbers in the different categories has commonly been done manually, machine learning algorithms offer an opportunity to automatise the process, while also deriving new outputs from existing datasets. This project used two newly developed image segmentation algorithms (YOLOv8 and Mask R-CNN) to extract features from X-ray images, and then used these features to count the number of seeds in images, and whether they were full, partially full, empty or infested (Duffy et al., 2024). To do so, the image was segmented into "outside", "inner cavity", "endosperm" and/or "infestation". These classes were chosen to encompass the variability in seed morphologies. This allowed the estimation of the percentage fullness of the seed to classify it as full, partially full or empty. It also allowed the identification of infested seeds. The next steps will include extracting morphological traits from these images, such as seed coat shape and thickness, and endosperm shape, which could be linked to genetic variation in the collections.

Case study 3: Seed traits for tree production.

This project aims to develop image-based seed sorting methods for native tree species to optimise their use in restoration projects and improve seedling production processes. To achieve this, image analysis and machine learning methods were applied to predict the germination of individual seeds (Gómez Barreiro et al., 2025). The study produced a comprehensive dataset at the seed level, including X-ray images, photographs and records on germination and embryo presence, along with extensive morphological descriptors extracted from these images. This dataset was then used to train machine learning models to predict seed germination and identify the most relevant traits for seed sorting. Beyond the initial goal of improving seed sorting for restoration, data from this project can also support conservation genetic research. Seeds used in this study have been collected and stored separately by maternal line in the Millennium Seed Bank, allowing for tracking each line throughout the tests. Other projects at the Royal Botanic Gardens, Kew, have been investigating the genetic diversity of populations using seeds from some of these same collections. Integrating genetic data with existing phenotypic and imaging datasets could reveal which morphological descriptors serve as effective proxies for genetic variation, thereby supporting trait-based approaches aimed at making the assessment of genetic diversity patterns scalable across multiple collections within a seedbank.

of subsets of individuals constituting a population (represented by N_C) defines how well the seed collection will genetically represent the provenance population. A small sample may result in a limited set of genetic variants or alleles and reduced effective population size (N_e) of any future population that will be established from that collection, decreasing its prospects of survival and resilience (e.g. Diaz-Martin et al., 2023). Additionally, when N_e of a seed collection is small, there is a high risk that the plants propagated from those seeds if used on their own, will lead to the establishment of an inbred population. In contrast, allelic redundancy, that is the occurrence of multiple copies of each allele that would buffer the loss of alleles due to physical loss of specimens, senescence or other stochastic events (Griffith et al., 2017; Hoban, Volk, et al., 2018; Schumacher et al., 2024) is achieved more easily in large metacollections, although it is more difficult to achieve for rare alleles. Therefore, seed collections sourced from a small number of maternal lines likely have less genetic diversity, on average, than those from larger pools and indicators based on minimum numbers may be suitable, and are certainly often tractable.

Although the N_e of a seed collection will be difficult to calculate, as explained below, very approximate estimates of N_e can be obtained based on the number of maternal lines. Assuming that each maternal line is unrelated, the number of maternal lines multiplied by the “rule of thumb” N_e/N_C ratio of 0.1 (as observed in many species, e.g., Frankham, 2021) might provide at least an approximate lower bound for N_e of a seed collection. However, the true N_e of the seed collection decreases if maternal lines are related, and the actual N_e is hard to calculate because the number of paternal lines contributing to the seed set is usually entirely unknown and even harder to approximate. It must be noted that the actual count of seeds in the collection (i.e. the number of seeds collected) can highly diverge from the N_e, as a seed collection sampled from few maternal lines is highly limited in its genetic diversity (Bragg et al., 2021; Rosenberger & Hoban, 2024; Zumwalde et al., 2022). The strong influence of the number of maternal lines on the N_e is one of the reasons why seed sampling guidance suggests maximising the number of maternal lines sampled (Griffith et al., 2019). The number of contributing paternal lines is also important, and to some degree can be predicted if the species' pollination biology and the sampled population density are known, and especially if prior genetic studies of parentage exist (e.g. Hoban, Kallow, & Trivedi, 2018). Although this is the case for a few species, it may be possible for species such as *Fraxinus* and *Quercus* spp. for which many genetic studies have been performed.

A body of work exists on the minimum sample size needed to conserve a predefined level of genetic diversity, most often 95% of the alleles existing in wild populations. The number of maternal lines needed per population to achieve this threshold has been estimated to be 30 for a fully outcrossing and 60 for a fully selfing species, so the common recommendation has been 50 maternal lines per population (Brown & Marshall, 1995). This minimum has been refined via several empirical and simulation studies. The number of maternal lines can be lower than 50 if (a) populations are not entirely isolated, such that some populations share some genetic material and assuming multiple populations are sampled, and/or (b) if many seeds per maternal plant are sampled, which can to some degree make up for reduced numbers of maternal lines, although, as noted above, sampling many seeds from few maternal lines is not a very effective approach (Hoban & Schlarbaum, 2014; Hoban & Strand, 2015; Rosenberger & Hoban, 2024).

The practical question is then, how many seeds should be collected? A number of studies of both rare and common plants suggest that the number varies according to plant biology, with more seeds needed for long-lived species, inbreeding or selfing species, species with short pollen and/or seed dispersal, species in which some individuals are not reproductive each year, and when the full spatial extent of the population cannot be sampled (Hoban et al., 2020; Hoban & Schlarbaum, 2014; Hoban & Strand, 2015; Kallow et al., 2021). More seeds may be needed to conserve SNPs as compared to microsatellite DNA variation, and more may be needed to conserve haplotype variation (Koontz et al., 2024; Reeves & Richards, 2017). On the other hand, conserving alleles with known adaptive function may actually require lower seed collection sizes than neutral diversity (Exposito-Alonso et al., 2022). For rare species, collection sizes of approximately 300 seeds, from dozens of maternal lines, from the existing known populations, may be sufficient to conserve 95% of the genetic variation, in the case studies examined so far (Hoban et al., 2020; Rosenberger et al., 2022), although sampling from dozens of maternal lines might not be possible in species that are very rare.

2.2 | Geographic data

Seedbank collections made from populations inhabiting a range of environmental conditions can capture adaptive variation which offers advantages in the specific environments where the populations occur. In such distinct environments, populations are subjected to varying intensities of selective pressures, which may lead to genetic divergence and the establishment of specific ecotypes (Kang et al., 2023; Kileeg et al., 2024). Capturing adaptive variation from multiple populations or ecotypes will make the collections more representative of the overall intraspecific diversity, and these collections might then be used for restoration interventions tailored to increasing resilience in specific environments. Such principles have allowed the establishment of eco-regional approaches to seed collecting, which in some cases have been validated using ecological and genetic data from a subset of species and common garden experiments. A specific example is the UK National Tree Seed Project (UKNTSP), a seed conservation project led by the Millennium Seed Bank, in which the sampling strategy was based on the UK Forestry Commission Native Seed Zones. The Native Seed Zones were established to guide appropriately adapted forest reproductive material (Herbert et al., 1999; Kallow & Trivedi, 2017) and are based on geo-physical boundaries. Additionally, in the UK species-specific zones for *Pinus sylvestris* were established based on analyses of biochemical variation among populations. A study aiming to validate the seed zones for this species used vegetation surveys and long-term climatic data (Whittet, 2017), and found that 26% of the sites were assigned to groups different from those expected based on biochemical variation; only four groups were found based on community composition. Examples of seed transfer zones for non-woody species are those established in Germany based on biogeographic boundaries, and in the United States (Pike et al., 2020). In Norway, a study attempting to identify seed transfer zones based on genetic variation in four species from alpine areas found that genetic groups were related to differences in reproductive strategies and phylogeographic factors (Jørgensen et al., 2016).

Quantitative metrics based on geographic coverage have been employed for numerous species in the "gap analysis" approach, possibly first presented by Hamann et al. (2005), further pioneered by Khoury and colleagues (Khoury et al., 2010; Khoury et al., 2015; Syfert et al., 2016, among others), and expanded by Beckman and colleagues in a series of studies (Beckman et al., 2019, 2021; Good et al., 2024) in different genera and more recently across very large numbers of plant species (e.g. thousands, Khoury et al., 2019, Khoury et al., 2020). This gap analysis can assign each species one of several scores, for geographic or ecological 'coverage'. Species can then be ranked according to those with the least eco-geographic space covered, and thus those that need the most additional sampling or protection (the approach is applicable to ex situ seed conservation or in situ protected areas). This has raised the alarm about the very large numbers of species with insufficient eco-geographic conservation coverage, and thus presumably insufficient genetic diversity conserved.

The simplest geographic data associated with seed collections that can be used for the purposes of inferring ecotypic variation and geographic coverage are location coordinates, which are normally associated with habitat and geographic descriptors. Data on climate, geology and soil type can also be populated in seed collection databases (or derived from coordinates and habitat descriptors, in case these were not recorded), allowing their linkage to known environmental stressors and providing a context for understanding adaptive traits such as drought and dry heat tolerance. For instance, plants from deserts and shrublands might accumulate drought- and/or dry heat-resistant traits at various life stages (seed to mature plants), that can be further validated by physiological and genomic studies (Gurvich et al., 2017; Visscher et al., 2021).

Georeferenced herbarium and seedbank data can be combined to develop models that not only predict current and future distributions but also establish a prioritisation framework for seed collections, considering factors such as species usefulness, extinction risk, storage behaviour and existing collection status (Hernandez et al., 2024). In addition, geographic data can guide simulations of how genetic variation differs across populations (e.g., Reeves & Richards, 2017). In Hoban, Kallow, and Trivedi (2018), geographic location data and the UK seed zones for *Fraxinus excelsior* populations were used to create a model of how genetic distribution might be spread across the UK. The UKNTSP collections for this species were then mapped against this to estimate how much of the genetic diversity (in terms of common and rare alleles) may have been conserved in the seed collection, and lastly to advise about improvements to seed sampling strategies.

2.3 | Phenotypic data

Seed collections can be complemented with measurements of phenotypic traits such as seed size and seed weight, among others. These measurements can be used to infer underlying genetic differences, which might be driven by environmental factors or evolutionary processes, and to prioritise populations for further studies of phenotypic and genotypic variation across different seed collections/locations (Di Santo et al., 2021). For example, at the species level, variance in seed size is estimated to be 50% genetic, which includes a large maternal component (Larios et al., 2023). Intraspecific diversity in seed size also shows evidence for a genetic component as variation in seed size among populations is correlated with dispersal mechanisms, which play a crucial role in driving evolutionary changes in seed and fruit traits (Zhang et al., 2020; Zhang & Wang, 2024).

Several traits relating to seed-covering structures have been assessed for their variation on a global scale. For instance, Visscher et al. (2025) measured the hardness of seed-covering structures of >400 woody species with orthodox seeds and found roles for morphology, dormancy, dispersal and precipitation in explaining part of the (>2000-fold) global variation in this trait. Furthermore, the ratio of seed covering structures to whole seed biomass has been measured for >1,300 species and was found to be primarily associated with seed desiccation response (Wu et al., 2024). Future studies could determine the level of intraspecific variation in these traits and to what degree this trait variation is reflective of genetic variation.

In addition to measurements in vivo, morphological traits can be measured using high-resolution images. New machine learning methods combined with the acquisition of digital images of seeds are opening new avenues to explore variation within seed collections at multiple levels (BOX 1).

2.4 | Physiological data

As for morphological data, physiological data might reveal genotypic differences among individuals within and between populations and observing such differences might help prioritise populations for further studies of intraspecific genetic diversity. For example, primary and secondary seed dormancy are traits that influence plant fitness through the timing of germination, and it is known that the depth of seed dormancy can vary between populations of wild species (Gomez-Cabellos et al., 2022; Kerdaffrec & Nordborg, 2017; Visscher et al., 2022). In *Arabidopsis*, primary seed dormancy depth has been shown to be highly heritable but also partly (e.g., 40% in Swedish populations) controlled by genotype interactions with the maternal environment (Kerdaffrec & Nordborg, 2017).

Carta et al. (2022) used the germination records of European seedbanks available in ENSCOBASE (the European Native Seed Conservation Consortium database; http://enscobase.maich.gr/germination.tml) to test the hypothesis that the germination niche of temperate plant species is shaped by their climatic requirements and phylogenetic relatedness. A similar approach could be used in future meta-analyses based on intraspecific variation in the germination niche, thus allowing the exploration of the relationship between such variation and intraspecific genetic diversity.

For non-dormant seeds, germination rate and timing are closely linked to environmental cues such as temperature and water availability. While variation amongst species is well reported for temperature (e.g., base, optimum and ceiling temperature) and water (e.g., base water potential) thresholds to germination (Maleki et al., 2024), the parental environment during seed production can fine-tune these traits (Fernández-Pascual et al., 2019), such that differences in germination exist within species, even between individuals of the same genetic background (Awan et al., 2018). Valencia-Gredilla et al. (2020) explored the intraspecific variation in the base temperature of germination in *Conyza bonariensis* and found three of four populations were significantly different (4.9, 6.9 and 8.9°C), while Rotundo et al. (2015) reported significant variation in the base water potential to achieve 50% germination across seven populations of *Poa lingularis* (ranging from −0.48 ± 0.02 to −1.17 ± 0.04 MPa). Other studies have shown that intraspecific variation in germination speed at different temperatures involves genetic variation in particular genes, such as *GF14h* in rice (Sugimura et al., 2024; Yoshida et al., 2022). Research is currently being carried out at the Royal Botanic Gardens, Kew, leveraging the vast seed collections of the UK flora available at the Millennium Seed Bank (Clubbe et al., 2020) to explore intraspecific variability in population genetics and thermal thresholds for seed germination of UK native species. The work focuses on those species which are either threatened or of ecological restoration interest (https://www.kew.org/science/our-science/projects/native-seeds-restoration-uk) with the aim of providing conservation management and ecological restoration recommendations (White et al., 2025).

2.5 | Molecular data

Molecular data deriving from physical specimens (also part of the data referred to as Digital Sequence Information – DSI, in policy terms) (Aubry, 2023) can provide direct information about the genetic diversity of the provenance populations. Characterising molecular variation has been achieved for vast collections of crop germplasm (Mascher et al., 2019), but the lack of such research on the less abundant wild-species germplasm collections also reflects the difficulty in prioritising populations that are representative of intraspecific genetic variation. In addition, to enable the characterisation of intraspecific genetic diversity in a collection for the purposes of assessing whether this is representative of the genetic diversity in the wild and how these should be used in restoration practices, molecular data should derive from the genetic analysis of several individuals within the provenance population (genotypic data). There are also more general considerations associated with access to and sharing of DSI and benefit sharing (Cowell et al., 2022), which include ensuring that correct attributions and permissions are in place (McCartney et al., 2022), so that any DSI made public can be can be traced back to its origin for benefit sharing were it to be used and that intellectual property is not claimed unethically by users (Anderson & Hudson, 2020).

3 | FUTURE PERSPECTIVES AND UNTAPPED POTENTIAL OF DIGITAL INFORMATION FOR CONSERVATION GENETICS

In this article, we showed how digital data associated with seedbanks can help increase our knowledge of intraspecific genetic diversity held in the collections, including the prioritisation of species for further genetic studies. Among the potential improvements that might increase the predictive power of digital data, we identify the possibility to unify or link databases, even if the immediate benefits are not evident. For example, unifying databases including field collection data, germination data, propagation data in nursery settings and plot data when seedlings are planted, might help trace the proportion of seeds that are lost at each stage and link this to data such as the number of mother trees and collection location, thus better-informing seed sourcing prior to restoration trials.

Among the challenges associated with the way seedbank digital data are currently collected and handled are those associated with the lack of knowledge, for instance when data are not recorded in the field, or for certain categories of species (e.g. those with recalcitrant seeds), and poor data processing and standardisation (Pedrini & Dixon, 2020). For the purposes of characterising genetic diversity in seedbank collections, the inclusion of data on life-history traits of the species and molecular data would allow the optimisation of collection strategies species-by-species.

An important challenge, also at the forefront of the current policy debate concerns access to data, as most databases are not publicly available. A notable exception is ENSCOBASE (http://enscobase.maich.gr/), an openly available database of seedbank digital information, which has been used to further conservation targets and research within Europe and is the collaborative effort of the European Native Seed Conservation Network (ENSCONET). Over 73,763 accessions are openly available, with data from 21,970 germination tests, and collections spanning the 11 biogeographical regions defined by the European Environmental Agency (EEA). This database has been used to assess progress towards conservation targets, such as the 2020 Global Strategy for Plant Conservation (GSPC) Targets 8 and 9 (Rivière et al., 2018), assessing the number of threatened plants and crop wild relatives conserved ex situ, their availability, and the level of intraspecific diversity conserved within these collections

(Rivière & Müller, 2017). Digitised seedbank data stored in large, open-access databases present an opportunity to link geographical data and germination data with conservation genetic studies to improve the precision of effective conservation assessments. A new Seed Portal for Online Data (SeedPOD) due to go live in 2026 will also address this lack of access to seed collection data. It will be the first open-access database of global wild-origin seed collections which links collections banked at multiple wild seedbanks. Detailed field, processing, seed quantity and test data will be available, plus a point of request for seed samples from all seedbanks sharing their data.

New opportunities include the possibility to find indicators for monitoring intraspecific genetic diversity in relation to policy targets, such as those of the Kunming-Montreal Global Biodiversity Framework (KMGBF) (https://www.cbd.int/gbf/targets). Among the KMGBF goals and targets, Target 4 specifically focuses on maintaining and restoring intraspecific genetic diversity, including through ex situ conservation practices. Similarly, GOAL A proposes to maintain the genetic diversity within populations of wild and domesticated species, safeguarding their adaptive potential. Towards these, the indicator "Number of plant and animal genetic resources secured in medium or long-term conservation facilities" is suggested as a component or complementary indicator, respectively. While this indicator can be easily measured leveraging seedbank records, such as the number of accessions, number of viable seeds per collection and number of mother plants, it does not necessarily provide information on intraspecific genetic diversity in the collection without considering other types of data. Among the possible data, can we find the most pragmatic proxies for intraspecific genetic variation in terms of feasibility and scalability? Which seed traits, if any, would be a good proxy for genetic variation generally? Given the fact that not many seed traits are characterised at intraspecific level for many species, which seed traits would be easiest to characterise intraspecifically at scale and within short time frames?

New opportunities are also offered by new technologies and featured in the projects described in Box 1. These projects highlight the value of integrating machine learning into conservation genetics pipelines and underscore the importance of digitising seed collections. As the largest wild-origin seedbank, the Millennium Seed Bank could become an exceptional hub for research in image-based conservation genetics if efforts were made to digitise its almost 100,000 collections (comprising more than 40,000 species). In recent years, significant investment has gone into digitising collections held by the Royal Botanic Gardens, Kew, [https://www.kew.org/science/digitising-kews-collections]. However, seeds held at the Millennium Seed Bank have not been included in these initiatives, and projects requiring image data from these seeds have instead been developed through small grants. Advances in high-throughput methods now allow the acquisition of extensive trait data and images from seeds in a time- and labour-saving manner (Dayrell et al., 2023). Automated image-based analysis offers a scalable means to generate large datasets of quantitative phenotypic information, facilitating the exploration of complex, multidimensional relationships between traits, their genetic bases, and environmental interactions (Lürig et al., 2021). Such high-dimensional data can reveal previously unrecognised axes of phenotypic variation and offer deeper insights into the mechanistic links between phenotype, genotype and environment. We hope our perspective will stimulate more work on these aspects.

AUTHOR CONTRIBUTIONS

RG designed the study with contributions from KAH, EB and PK. RG coordinated author contributions and led the writing. SH, AMV, PGB, RLCD, KLDA, ARH, CES, SK, JS, LO, NC, TC, EM, PK, EB and KAH wrote the manuscript, with contributions from CCa, CCo, MFF, JV, IB, SM, TP.

ACKNOWLEDGEMENTS

We thank the editor and two reviewers for their constructive comments on the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No new data were generated during the preparation of this article.

ORCID

Roberta Gargiulo https://orcid.org/0000-0001-8663-6568
Sean Hoban https://orcid.org/0000-0002-0348-8449
Anne M. Visscher https://orcid.org/0000-0002-7005-9142
Pablo Gómez Barreiro https://orcid.org/0000-0002-3140-3326
Roberta L. C. Dayrell https://orcid.org/0000-0002-4770-9100
Alice R. Hudson https://orcid.org/0000-0003-3860-9199
Charlotte E. Seal https://orcid.org/0000-0002-9329-9325
Simon Kallow https://orcid.org/0000-0002-0869-5797
Jan Sala https://orcid.org/0000-0002-8896-1265
Efisio Mattana https://orcid.org/0000-0001-6235-4603
Clare Callow https://orcid.org/0009-0007-0944-7617
Christopher Cockel https://orcid.org/0000-0002-7589-3910
Michael F. Fay https://orcid.org/0000-0003-3491-9093
Stephanie Miles https://orcid.org/0000-0002-1478-0808
Timothy Pearce https://orcid.org/0000-0003-1892-6824
Paul Kersey https://orcid.org/0000-0002-7054-800X
Elinor Breman https://orcid.org/0000-0001-9834-5186
Kate A. Hardwick https://orcid.org/0000-0001-7864-8008

REFERENCES

Anderson, J., & Hudson, M. (2020). The biocultural labels initiative: Supporting indigenous rights in data derived from genetic resources. *Biodiversity Information Science and Standards*, *4*, e59230. https://doi.org/10.3897/biss.4.59230

Aubry, S. (2023). Genebanking plant genetic resources in the postgenomic era. *Agriculture and Human Values*, *40*(3), 961–971. https://doi.org/10.1007/s10460-023-10417-7

Awan, S., Footitt, S., & Finch-Savage, W. E. (2018). Interaction of maternal environment and allelic differences in seed vigour genes determines seed performance in *Brassica oleracea*. *The Plant Journal*, *94*(6), 1098–1108. https://doi.org/10.1111/tpj.13922

Beckman, E., Meyer, A., Denvir, A., Gill, D., Man, G., Pivorunas, D., Shaw, K., & Westwood, M. (2019). Conservation Gap Analysis of Native U.S. Oaks. Lisle, IL: The Morton Arboretum. Accessed March 11, 2025. https://mortonarb.org/app/uploads/2021/05/conservation-gap-analysis-of-native-US-oaks_sm.pdf

Beckman, E., Meyer, A., Pivorunas, D., Hoban, S., & Westwood, M. (2021). Conservation Gap Analysis of Native U. S Yews. https://www.mortonarb.org/app/uploads/2021/05/conservation_gap_analysis_of_native_us_yews.pdf

Borrell, J. S., Zohren, J., Nichols, R. A., & Buggs, R. J. (2020). Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. *Evolutionary Applications*, *13*(1), 161–175. https://doi.org/10.1111/eva.12883

Bragg, J. G., Yap, J. Y. S., Wilson, T., Lee, E., & Rossetto, M. (2021). Conserving the genetic diversity of condemned populations: Optimizing collections and translocation. *Evolutionary Applications*, *14*(5), 1225–1238. https://doi.org/10.1111/eva.13192

Brown, A. H. D., & Hardner, C. M. (2000). Sampling the gene pools of forest trees for ex situ conservation. In A. Young, D. Boshier, & T. Boyle (Eds.), *Forest conservation genetics: Principles and practice* (pp. 185–196). CAB International. https://doi.org/10.1079/9780851995045.0185

Brown, A. H. D., & Marshall, D. R. (1995). A basic sampling strategy: theory and practice. In L. Guarino, V. Ramanatha Rao, & R. Reid (Eds.), *Collecting plant genetic diversity: Technical guidelines* (pp. 75–91). CAB International.

Bucharova, A., Bossdorf, O., Hölzel, N., Kollmann, J., Prasse, R., & Durka, W. (2019). Mix and match: Regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. *Conservation Genetics*, *20*(1), 7–17. https://doi.org/10.1007/s10592-018-1067-6

Burbano, H. A., & Gutaker, R. M. (2023). Ancient DNA genomics and the renaissance of herbaria. *Science*, *382*(6666), 59–63. https://doi.org/10.1126/science.adi1180

Carta, A., Fernández-Pascual, E., Gioria, M., Müller, J. V., Rivière, S., Rosbakh, S., Saatkamp, A., Vandelook, F., & Mattana, E. (2022). Climate shapes the seed germination niche of temperate flowering plants: A meta-analysis of European seed conservation data. *Annals of Botany*, *129*(7), 775–786. https://doi.org/10.1093/aob/mcac037

Cheng, S., Feng, C., Wingen, L. U., Cheng, H., Riche, A. B., Jiang, M., & Griffiths, S. (2024). Harnessing landrace diversity empowers wheat breeding. *Nature*, *632*(8026), 823–831. https://doi.org/10.1038/s41586-024-07682-9

Clubbe, C., Ainsworth, A. M., Bárrios, S., Bensusan, K., Brodie, J., Cannon, P., et al. (2020). Current knowledge, status, and future for plant and fungal diversity in Great Britain and the UK overseas territories. *Plants*, *People*, *Planet*, *2*(5), 557–579. https://doi.org/10.1002/ppp3.10142

Cowell, C., Paton, A., Borrell, J. S., Williams, C., Wilkin, P., Antonelli, A., & Kersey, P. J. (2022). Uses and benefits of digital sequence information from plant genetic resources: Lessons learnt from botanical collections. *Plants*, *People*, *Planet*, *4*(1), 33–43. https://doi.org/10.1002/ppp3.10216

Dayrell, R. L. C., Ott, T., Horrocks, T., & Poschlod, P. (2023). Automated extraction of seed morphological traits from images. *Methods in Ecology and Evolution*, *14*, 1708–1718. https://doi.org/10.1111/2041-210X.14127

Di Sacco, A., Hardwick, K. A., Blakesley, D., Brancalion, P. H., Breman, E., Cecilio Rebola, L., … Antonelli, A. (2021). Ten golden rules for reforestation to optimize carbon sequestration, biodiversity recovery and livelihood benefits. *Global Change Biology*, *27*(7), 1328–1348. https://doi.org/10.1111/gcb.15498

Di Santo, L. N., Polgar, M., Nies, S., Hodgkiss, P., Canning, C. A., Wright, J. W., & Hamilton, J. A. (2021). Seed morphological traits as a tool to quantify variation maintained in ex situ collections: A case study in Pinus torreyana. *AoB Plants*, *13*(5), plab058. https://doi.org/10.1093/aobpla/plab058

Diaz-Martin, Z., Fant, J., Havens, K., Cinea, W., Lima, J. M. T., & Griffith, M. P. (2023). Current management practices do not adequately safeguard endangered plant species in conservation collections. *Biological Conservation*, *280*, 109955. https://doi.org/10.1016/j.biocon.2023.109955

Duffy, G., Gyasi, V., Holmes, A., Morton, A., Roberts, C., Rosen, O., Stoyanov, I., Craven, R., & Dhanjal-Adams, K. L. (2024). *Identifying Viable Seeds from X-Rays*. MSc thesis. Imperial College London.

Edet, O. U., Gorafi, Y. S., Nasuda, S., & Tsujimoto, H. (2018). DArTseq-based analysis of genomic relationships among species of tribe Triticeae. *Scientific Reports*, *8*(1), 16397.

Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., … Zess, E. (2022). Genetic diversity loss in the Anthropocene. *Science*, *377*(6613), 1431–1435. https://doi.org/10.1126/science.abn5642

Fernández-Pascual, E., Mattana, E., & Pritchard, H. W. (2019). Seeds of future past: Climate change and the thermal memory of plant reproductive traits. *Biological Reviews*, *94*(2), 439–456. https://doi.org/10.1111/brv.12461

Frankham, R. (2021). Suggested improvements to proposed genetic indicator for CBD. *Conservation Genetics*, *22*(4), 531–532. https://doi.org/10.1007/s10592-021-01357-y

Franks, S. J., Hamann, E., & Weis, A. E. (2018). Using the resurrection approach to understand contemporary evolution in changing environments. *Evolutionary Applications*, *11*(1), 17–28. https://doi.org/10.1111/eva.12528

Gargiulo, R., Budde, K. B., & Heuertz, M. (2024). Mind the lag: Understanding genetic extinction debt for conservation. *Trends in Ecology & Evolution*, *40*, 228–237. https://doi.org/10.1016/j.tree.2024.10.008

Gargiulo, R., Saubin, M., Rizzuto, G., West, B., Fay, M. F., Kallow, S., & Trivedi, C. (2019). Genetic diversity in British populations of Taxus baccata L.: Is the seedbank collection representative of the genetic variation in the wild? *Biological Conservation*, *233*, 289–297. https://doi.org/10.1016/j.biocon.2019.01.014

Gómez Barreiro, P., Richard-Bollans, A., Mattana, E., Seal, C. E., Chapman, T., & Dayrell, R. (2025). Enhancing tree seed germination prediction with image-driven machine learning models. *bioRxiv*, 2025–01. https://doi.org/10.1101/2025.01.21.634084

Gomez-Cabellos, S., Toorop, P. E., Cañal, M. J., Iannetta, P. P. M., Fernández-Pascual, E., Pritchard, H. W., & Visscher, A. M. (2022). Global DNA methylation and cellular 5-methylcytosine and H4 acetylated patterns in primary and secondary dormant seeds of *Capsella bursa-pastoris* (L.) Medik. (shepherd's purse). *Protoplasma*, *259*, 595–614. https://doi.org/10.1007/s00709-021-01678-2

Good, K., Rivera Huerta, H., Luna, L., Serrano, J., Garcia Balderas, V., and Alvarez-Clare, S. (2024). Quercus cedrosensis C.H.Mull. In Good, K., Coombes, A. J., Valencia-A, S., Rodríguez-Acosta, M., Beckman Bruns, E., and Alvarez-Clare, S. *Conservation gap analysis of native Mesoamerican oaks* (pp. 101–108). The Morton Arboretum.

Goodale, U. M., Antonelli, A., Nelson, C. R., & Chau, M. M. (2023). Seed banks needed to restore ecosystems. *Science*, *379*(6628), 147.

Gosling, P. G. (2003). Viability testing. In Smith, R. D., Dickie, J. D., Linnington, S. H., Pritchard, H. W. & Probert, R. J. *Seed conservation: Turning science into practice*. RBG Kew.

Griffith, M. P., Beckman, E., Calicrate, T., Clark, J. R., Clase, T., Deans, S., & Wood, J. (2019). *Toward the metacollection: Safeguarding plant diversity and coordinating conservation collections*. Botanic Gardens

Conservation International-US). Link: Toward the Metacollection | Botanic Gardens Conservation International Accessed March 11 2025.

Griffith, M. P., Calonje, M., Meerow, A. W., Francisco-Ortega, J., Knowles, L., Aguilar, R., … Magellan, T. M. (2017). Will the same ex situ protocols give similar results for closely related species? *Biodiversity and Conservation*, *26*, 2951–2966. https://doi.org/10.1007/s10531-017-1400-2

Gurvich, D. E., Pérez-Sánchez, R., Bauk, K., Jurado, E., Ferrero, M. C., Funes, G., & Flores, J. (2017). Combined effect of water potential and temperature on seed germination and seedling development of cacti from a Mesic argentine ecosystem. *Flora*, *227*, 18–24. https://doi.org/10.1016/j.flora.2016.12.003

Halewood, M., Chiurugwi, T., Sackville Hamilton, R., Kurtz, B., Marden, E., Welch, E., … Powell, W. (2018). Plant genetic resources for food and agriculture: Opportunities and challenges emerging from the science and information technology revolution. *New Phytologist*, *217*(4), 1407–1419. https://doi.org/10.1111/nph.14993

Hamann, A., Smets, P., Yanchuk, A. D., & Aitken, S. N. (2005). An ecogeographic framework for in situ conservation of forest trees in British Columbia. *Canadian Journal of Forest Research*, *35*(11), 2553–2561. https://doi.org/10.1139/x05-181

Herbert, R., Samuel, S., & Patterson, G. (1999). Using local stock for planting native trees and shrubs. Accessed March 11, 2025. https://cdn.forestresearch.gov.uk/2022/02/fcpn8-1.pdf

Hernandez, M. G. C., Gómez Barreiro, P., White, J. D. M., & Way, M. (2024). Sonoran Desert ex situ conservation gap analysis: charting the path towards conservation. *bioRxiv*, 2024–11.

Hoban, S., Bruford, M. W., da Silva, J. M., Funk, W. C., Frankham, R., Gill, M. J., & Laikre, L. (2023). Genetic diversity goals and targets have improved, but remain insufficient for clear implementation of the post-2020 global biodiversity framework. *Conservation Genetics*, *24*(2), 181–191. https://doi.org/10.1007/s10592-022-01492-0

Hoban, S., Callicrate, T., Clark, J., Deans, S., Dosmann, M., Fant, J., & Griffith, M. P. (2020). Taxonomic similarity does not predict necessary sample size for ex situ conservation: A comparison among five genera. *Proceedings of the Royal Society B*, *287*(1926), 20200102. https://doi.org/10.1098/rspb.2020.0102

Hoban, S., Kallow, S., & Trivedi, C. (2018). Implementing a new approach to effective conservation of genetic diversity, with ash (Fraxinus excelsior) in the UK as a case study. *Biological Conservation*, *225*, 10–21. https://doi.org/10.1016/j.biocon.2018.06.017

Hoban, S., & Schlarbaum, S. (2014). Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. *Biological Conservation*, *177*, 90–99. https://doi.org/10.1016/j.biocon.2014.06.014

Hoban, S., & Strand, A. (2015). Ex situ seed collections will benefit from considering spatial sampling design and species' reproductive biology. *Biological Conservation*, *187*, 182–191. https://doi.org/10.1016/j.biocon.2015.04.023

Hoban, S., Volk, G., Routson, K. J., Walters, C., & Richards, C. (2018). Sampling wild species to conserve genetic diversity. In *North American crop wild relatives*, volume 1: Conservation strategies (pp. 209–228). Springer.

Jørgensen, M. H., Elameen, A., Hofman, N., Klemsdal, S., Malaval, S., & Fjellheim, S. (2016). What's the meaning of local? Using molecular markers to define seed transfer zones for ecological restoration in Norway. *Evolutionary Applications*, *9*(5), 673–684. https://doi.org/10.1111/eva.12378

Kallow, S., Panis, B., Vu, D. T., Vu, T. D., Paofa, J., Mertens, A., & Janssens, S. B. (2021). Maximizing genetic representation in seed collections from populations of self and cross-pollinated banana wild relatives. *BMC Plant Biology*, *21*, 1–14.

Kallow, S., Trivedi, C. (2017). Collecting genetic variation on a small island. In: Sniezko, Richard A.; Man, Gary; Hipkins, Valerie; Woeste, Keith; Gwaze, David; Kliejunas, John T.; McTeague, Brianna A., tech. cords. 2017. Gene conservation of tree species—banking on the future. Proceedings of a workshop. Gen. Tech. Rep. PNW-GTR-963. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: 129–136.

Kang, M., Wu, H., Liu, H., Liu, W., Zhu, M., Han, Y., Liu, W., Chen, C., Song, Y., Tan, L., Yin, K., Zhao, Y., Yan, Z., Lou, S., Zan, Y., & Liu, J. (2023). The pan-genome and local adaptation of *Arabidopsis thaliana*. *Nature Communications*, *14*, 6259. https://doi.org/10.1038/s41467-023-42029-4

Kerdaffrec, E., & Nordborg, M. (2017). The maternal environment interacts with genetic variation in regulating seed dormancy in Swedish *Arabidopsis thaliana*. *PLoS ONE*, *12*, e0190242. https://doi.org/10.1371/journal.pone.0190242

Khoury, C., Laliberté, B., & Guarino, L. (2010). Trends in ex situ conservation of plant genetic resources: A review of global crop and regional conservation strategies. *Genetic Resources and Crop Evolution*, *57*, 625–639. https://doi.org/10.1007/s10722-010-9534-z

Khoury, C. K., Amariles, D., Soto, J. S., Diaz, M. V., Sotelo, S., Sosa, C. C., & Jarvis, A. (2019). Comprehensiveness of conservation of useful wild plants: An operational indicator for biodiversity and sustainable development targets. *Ecological Indicators*, *98*, 420–429. https://doi.org/10.1016/j.ecolind.2018.11.016

Khoury, C. K., Carver, D., Greene, S. L., Williams, K. A., Achicanoy, H. A., Schori, M., … Frances, A. (2020). Crop wild relatives of the United States require urgent conservation action. *Proceedings of the National Academy of Sciences*, *117*(52), 33351–33357. https://doi.org/10.1073/pnas.2007029117

Khoury, C. K., Heider, B., Castañeda-Álvarez, N. P., Achicanoy, H. A., Sosa, C. C., Miller, R. E., & Struik, P. C. (2015). Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) lam., I. Series Batatas]. *Frontiers in Plant Science*, *6*, 251.

Kileeg, Z., Wang, P., & Mott, G. A. (2024). Chromosome-scale assembly and annotation of eight Arabidopsis thaliana ecotypes. *Genome Biology and Evolution*, *16*, 1–8. https://doi.org/10.1093/gbe/evae169

Koontz, A. C., Salas, G. D., & Hoban, S. M. (2024). Genetic marker type impacts ex situ conservation minimum sample size estimates and their variance. https://doi.org/10.21203/rs.3.rs-5356916/v1

Larios, E., Ramirez-Parada, T. H., & Mazer, S. J. (2023). Heritability and variance components of seed size in wild species: Influences of breeding design and the number of genotypes tested. *Heredity*, *130*, 251–258. https://doi.org/10.1038/s41437-023-00597-7

Li, D. Z., & Pritchard, H. W. (2009). The science and economics of ex situ plant conservation. *Trends in Plant Science*, *14*(11), 614–621. https://doi.org/10.1016/j.tplants.2009.09.005

Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A., & Tsuboi, M. (2021). Computer vision, machine learning, and the promise of Phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution*, *9*, 642774. https://doi.org/10.3389/fevo.2021.642774

Maleki, K., Soltani, E., Seal, C. E., Colville, L., Pritchard, H. W., & Lamichhane, J. R. (2024). The seed germination spectrum of 486 plant species: A global meta-regression and phylogenetic pattern in relation to temperature and water potential. *Agricultural and Forest Meteorology*, *346*, 109865. https://doi.org/10.1016/j.agrformet.2023.109865

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., & Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nature Genetics*, *51*(7), 1076–1081. https://doi.org/10.1038/s41588-019-0443-6

Mattana, E., Ulian, T., & Pritchard, H. W. (2022). Seeds as natural capital. *Trends in Plant Science*, *27*(2), 139–146. https://doi.org/10.1016/j.tplants.2021.08.008

McCartney, A. M., Anderson, J., Liggins, L., Hudson, M. L., Anderson, M. Z., TeAika, B., & Phillippy, A. M. (2022). Balancing openness with indigenous data sovereignty: An opportunity to leave no one behind in the journey to sequence all of life. *Proceedings of the National Academy of Sciences*, *119*(4), e2115860119. https://doi.org/10.1073/pnas.2115860119

McGlaughlin, M. E., Riley, L., Brandsrud, M., Arcibal, E., Helenurm, M. K., & Helenurm, K. (2015). How much is enough? Minimum sampling intensity required to capture extant genetic diversity in ex situ seed collections: Examples from the endangered plant *Sibara filifolia* (Brassicaceae). *Conservation Genetics*, *16*, 253–266. https://doi.org/10.1007/s10592-014-0655-3

Milner, S. G., Jost, M., Taketa, S., Mazón, E. R., Himmelbach, A., Oppermann, M., & Stein, N. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, *51*(2), 319–326. https://doi.org/10.1038/s41588-018-0266-x

Namoff, S., Husby, C. E., Francisco-Ortega, J., Noblick, L. R., Lewis, C. E., & Griffith, M. P. (2010). How well does a botanical garden collection of a rare palm capture the genetic variation in a wild population? *Biological Conservation*, *143*(5), 1110–1117. https://doi.org/10.1016/j.biocon.2010.02.004

Pedrini, S., & Dixon, K. W. (2020). International principles and standards for native seeds in ecological restoration. *Restoration Ecology*, *28*, S286–S303. https://doi.org/10.1111/rec.13155

Pike, C., Potter, K. M., Berrang, P., Crane, B., Baggs, J., Leites, L., & Luther, T. (2020). New seed-collection zones for the eastern United States: The eastern seed zone forum. *Journal of Forestry*, *118*(4), 444–451. https://doi.org/10.1093/jofore/fvaa013

Pradham, N., Fan, X., Martini, F., Chen, H., Liu, H., Gao, J., & Manage Goodale, U. (2022). Seed viability testing for research and conservation of epiphytic and terrestrial orchids. *Botanical Studies*, *63*, 3. https://doi.org/10.1186/s40529-022-00333-0

Reeves, P. A., & Richards, C. M. (2017). Capturing haplotypes in germplasm core collections using bioinformatics. *Genetic Resources and Crop Evolution*, *64*, 1821–1828. https://doi.org/10.1007/s10722-017-0549-6

Richards, C. M., Lockwood, D. R., Volk, G. M., & Walters, C. (2010). Modeling demographics and genetic diversity in ex situ collections during seed storage and regeneration. *Crop Science*, *50*(6), 2440–2447. https://doi.org/10.2135/cropsci2010.04.0236

Rivière, S., Breman, E., Kiehn, M., Carta, A., & Müller, J. V. (2018). How to meet the 2020 GSPC target 8 in Europe: Priority-setting for seed banking of native threatened plants. *Biodiversity and Conservation*, *27*(8), 1873–1890. https://doi.org/10.1007/s10531-018-1513-2

Rivière, S., & Müller, J. V. (2017). Contribution of seed banks across Europe towards the 2020 global strategy for plant conservation targets, assessed through the ENSCONET database. *Oryx*, *52*(3), 464–470. https://doi.org/10.1017/S0030605316001496

Rosenberger, K., Schumacher, E., Brown, A., & Hoban, S. (2022). Species-tailored sampling guidelines remain an efficient method to conserve genetic diversity ex situ: A study on threatened oaks. *Biological Conservation*, *275*, 109755. https://doi.org/10.1016/j.biocon.2022.109755

Rosenberger, K. J., & Hoban, S. (2024). Simulating pollen flow and field sampling constraints helps revise seed sampling recommendations for conserving genetic diversity. *Applications in Plant Sciences*, *12*, e11561.

Rotundo, J. L., Aguiar, M. R., & Benech-Arnold, R. (2015). Understanding erratic seedling emergence in perennial grasses using physiological models and field experimentation. *Plant Ecology*, *216*, 143–156. https://doi.org/10.1007/s11258-014-0423-y

Schumacher, E., Wu, Y., Byrne, A., Gray, S., Ladd, L., Griffith, M. P., & Hoban, S. (2024). Examining previously neglected aspects of ex situ gene conservation in two IUCN Threatened plant species: Rare alleles, redundancy, ecogeographic representativeness, and relatedness. *International Journal of Plant Sciences*, *185*(2), 198–209.

Sugimura, Y., Oikawa, K., Sugihara, Y., Utsushi, H., Kanzaki, E., Ito, K., Ogasawara, Y., Fujioka, T., Takagi, H., Shimizu, M., Shimono, H., Terauchi, R., & Abe, A. (2024). Impact of rice *GENERAL REGULATORY FACTOR14h* (*GF14h*) on low-temperature seed germination and its application to breeding. *PLoS Genetics*, *20*, e1011369. https://doi.org/10.1371/journal.pgen.1011369

Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., & Chen, F. (2020). The wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnology Journal*, *18*(6), 1354–1360. https://doi.org/10.1111/pbi.13361

Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., … Knapp, S. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): Poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany*, *103*(4), 635–651. https://doi.org/10.3732/ajb.1500539

Valencia-Gredilla, F., Supiciche, M. L., Chantre, G. R., Recasens, J., & Royo-Esnal, A. (2020). Germination behaviour of *Conyza bonariensis* to constant and alternating temperatures across different populations. *Annals of Applied Biology*, *176*, 36–46. https://doi.org/10.1111/aab.12556

Visscher, A. M., Boatfield, M., Klak, C., Yeo, M., Pearce, T. R., Wilman, V., Mdayi, N., Gumede, S., & Pritchard, H. W. (2022). Physiological seed dormancy of *Ruschia imbricata* and *Ruschia uitenhagensis* (Aizoaceae) is broken by dry heat and unaffected by seasonality. *South African Journal of Botany*, *147*, 457–466. https://doi.org/10.1016/j.sajb.2022.02.007

Visscher, A. M., Gómez Barreiro, P., Soto Gomez, M., Carta, A., Liu, U., Wu, Y., & Pritchard, H. W. (2025). Global variation in seed covering structure hardness of woody species with orthodox seeds. *Annals of Botany*, mcaf027. https://doi.org/10.1093/aob/mcaf027

Visscher, A. M., Latorre Frances, A., Yeo, M., Yan, J., Colville, L., Gomez Barreiro, P., & Pritchard, H. W. (2021). Comparative analyses of extreme dry seed thermotolerance in five Cactaceae species. *Environmental and Experimental Botany*, *188*, 104514. https://doi.org/10.1016/j.envexpbot.2021.104514

Volk, G. M., Richards, C. M., Reilley, A. A., Henk, A. D., Forsline, P. L., & Aldwinckle, H. S. (2005). Ex situ conservation of vegetatively propagated species: Development of a seed-based core collection for Malus sieversii. *Journal of the American Society for Horticultural Science*, *130*(2), 203–210. https://doi.org/10.21273/JASHS.130.2.203

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., & Leung, H. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, *557*(7703), 43–49. https://doi.org/10.1038/s41586-018-0063-9

White, K., Cornwell-Davison, F., Cockel, C., Chapman, T., Mattana, E., & Viruel, J. (2025). Application of a biological trait-based framework for plant species conservation assessments in ecological restoration. *Restoration Ecology*, *33*, e14391. https://doi.org/10.1111/rec.14391

Whittet, R. (2017). Genetic resources of native tree species and their deployment under climate change. PhD Thesis. University of Edinburgh.

Wu, L.-M., Chen, S.-C., Quan, R.-C., & Wang, B. (2024). Disentangling the relative contributions of factors determining seed physical defence: A global-scale data synthesis. *Functional Ecology*, *38*, 1146–1155. https://doi.org/10.1111/1365–2435.14552

Yoshida, H., Hirano, K., Yano, K., Wang, F., Mori, M., Kawamura, M., Koketsu, E., Hattori, M., Ordonio, R. L., Huang, P., Yamamoto, E., & Matsuoka, M. (2022). Genome-wide association study identifies a gene responsible for temperature-dependent rice germination. *Nature Communications*, *13*, 5665. https://doi.org/10.1038/s41467-022-33318-5

Zhang, C., Li, J., Willis, C. G., & Ma, Z. (2020). Among-population variation in seed mass for 190 Tibetan plant species: Phylogenetic pattern and ecological correlates. *Global Ecology and Conservation*, *23*, e01163. https://doi.org/10.1016/j.gecco.2020.e01163

Zhang, J., & Wang, B. (2024). Intraspecific variation in seed size is mediated by seed dispersal modes and animal dispersers – Evidence from a global-scale dataset. *New Phytologist*, *241*, 461–470. https://doi.org/10.1111/nph.19340

Zumwalde, B. A., Fredlock, B., Beckman Bruns, E., Duckett, D., McCauley, R. A., Spence, E. S., & Hoban, S. (2022). Assessing ex situ genetic and ecogeographic conservation in a threatened but widespread oak after range-wide collecting effort. *Evolutionary Applications*, *15*(6), 1002–1017. https://doi.org/10.1111/eva.13391

How to cite this article: Gargiulo, R., Hoban, S., Visscher, A. M., Gómez Barreiro, P., Dayrell, R. L. C., Dhanjal-Adams, K. L., Hudson, A. R., Seal, C. E., Kallow, S., Sala, J., Oliver, L., Carvey, N., Chapman, T., Mattana, E., Callow, C., Cockel, C., Fay, M. F., Viruel, J., Birchenko, I., … Hardwick, K. A. (2025). The potential of seedbank digital information in plant conservation. *Plants*, *People*, *Planet*, 1–12. https://doi.org/10.1002/ppp3.70017