

Article

Capturing Genetic Diversity in Seed Collections: An Empirical Study of Two Congeners with Contrasting Mating Systems

Patricia Lu-Irving ¹, Jason G. Bragg ¹, Maurizio Rossetto ¹, Kit King ¹, Mitchell O'Brien ^{1,2}
and Marlien M. van der Merwe ^{1,*}

¹ Research Centre for Ecosystem Resilience, Australian Institute of Botanical Science, Royal Botanic Gardens Sydney, Mrs Macquaries Rd., Sydney, NSW 2000, Australia

² Commonwealth Scientific and Industrial Research Organisation (CSIRO), Innovation Quarter Westmead, Level 3, East Tower, 158-164 Hawkesbury Rd., Westmead, NSW 2145, Australia

* Correspondence: marlien.vandermerwe@botanicgardens.nsw.gov.au

Abstract: Plant mating systems shape patterns of genetic diversity and impact the long-term success of populations. As such, they are relevant to the design of seed collections aiming to maximise genetic diversity (e.g., germplasm conservation, ecological restoration). However, for most species, little is known empirically about how variation in mating systems and genetic diversity is distributed. We investigated the relationship between genetic diversity and mating systems in two functionally similar, co-occurring species of *Hakea* (Proteaceae), and evaluated the extent to which genetic diversity was captured in seeds. We genotyped hundreds of seedlings and mother plants via DArTseq, and developed novel implementations of two approaches to inferring the mating system from SNP data. A striking contrast in patterns of genetic diversity between *H. sericea* and *H. teretifolia* was revealed, consistent with a contrast in their mating systems. While both species had mixed mating systems, *H. sericea* was found to be habitually selfing, while *H. teretifolia* more evenly employed both selfing and outcrossing. In both species, seed collection schemes maximised genetic diversity by increasing the number of maternal lines and sites sampled, but twice as many sites were needed for the selfing species to capture equivalent levels of genetic variation at a regional scale.

Keywords: germplasm; *Hakea*; inbreeding; outcrossing; selfing



Citation: Lu-Irving, P.; Bragg, J.G.; Rossetto, M.; King, K.; O'Brien, M.; van der Merwe, M.M. Capturing Genetic Diversity in Seed Collections: An Empirical Study of Two Congeners with Contrasting Mating Systems. *Plants* **2023**, *12*, 522. <https://doi.org/10.3390/plants12030522>

Academic Editor: Calvin O. Qualset

Received: 10 November 2022

Revised: 10 January 2023

Accepted: 13 January 2023

Published: 23 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In plant conservation and restoration, maximising genetic diversity in seed collections aims to support the persistence of re-established populations by promoting their adaptive potential [1–3]. In addition to enabling adaptive evolution in the long term, adequate genetic diversity can mitigate inbreeding depression as well as determine reproductive success in the short term, i.e., in outcrossing and self-incompatible species [4–6]. Plant mating systems shape patterns of genetic diversity among and within species and populations [7–9] but are often overlooked in the design of seed collections because they are poorly understood in most species.

The mating system is likely to be a major determinant of the extent to which seed collected under different strategies reflects the genetic diversity of standing populations. When mating systems are more selfing (vs. more outcrossing), it has been suggested that more effort is required to reliably capture genetic diversity both within populations [10] and among them (due to patterns in population structure; [11,12]). However, specific advice is sometimes lacking, particularly regarding the number of populations to sample to capture species-level genetic diversity [13], with general recommendations ranging from 5 sites to 50 [14–17] in order to meet the benchmark of a 95% probability of capturing all common alleles (>5% frequency) [10]. Empirical tests of the effect of the mating system on genetic diversity represented in seed collections are relatively few (notable examples include [18,19]).

While landscape-scale genomic data are increasingly applied to inform ecological restoration and conservation practice [20,21], there are few analytical tools for inferring the mating system from these data, and thus limited evidence of how variation in genetic diversity and mating systems are interrelated at this scale. In this study, we used recently described approaches [22] to characterise landscape-scale patterns in population genetics estimates, and developed new methods to infer patterns in mating systems. We tested whether similar species share similar patterns by comparing results from two congeneric, co-occurring, and ecologically similar species. Our goal was to provide empirical evidence to substantiate existing germplasm collection guidelines, and to support potential improvements to them. Specifically, we asked how the mating system might influence the genetic diversity captured by different approaches to seed collecting, and whether this might differ between the two species.

Data from genome-wide markers currently provides the most valuable proxy for understanding and capturing genetic variation for conservation purposes [23,24]. We genotyped thousands of Single Nucleotide Polymorphisms (SNPs) in hundreds of individuals of *Hakea sericea* and *Hakea teretifolia*, common shrubs of temperate Eastern Australia. We took a two-tiered approach to sampling. First, we sampled adult populations throughout the distributions of the species to characterise patterns of genetic diversity in the landscape. Then, we sampled wild plants and their progeny at a subset of sites to explore relationships between the genetic diversity, mating system, and the genetic diversity captured using different seed collecting approaches within and among sites. Tools to infer the mating system in incompletely sampled families from large SNP datasets are still emerging, so we used two approaches for estimating selfing rates from mother–seedling progeny arrays. The first was a reimplement of Ritland's [25] method of moments estimator (MME hereafter) for inferring individual single-locus selfing rates. The second is a novel approach to estimating individual levels of inbreeding based on a Convolutional Neural Network model (CNN hereafter). We compared selfing rates with patterns in heterozygosity, F-statistics, and relatedness, evaluating patterns in the context of seed collections for conservation and restoration.

2. Results

2.1. Genetic Diversity

2.1.1. Landscape Genomic Patterns

Analyses of broad-scale landscape genomic data revealed contrasting patterns between the two congeneric, co-distributed (Figure 1), and morphologically similar species. With a similar sampling strategy and sequence density, the DArTseq method captured SNP matrices of different sizes with 31,031 quality-filtered SNPs (from 48,728 raw SNPs) for *H. sericea* (203 total individuals across 31 sites) and 40,342 SNPs (from 80,527) for *H. teretifolia* (206 total individuals across 32 sites). Estimates of F_{ST} were higher in *H. sericea* (average 0.73) than in *H. teretifolia* (average 0.30), and estimates of observed (H_o) and expected heterozygosity (H_e) were lower in *H. sericea* (H_o average = 0.024; H_e average = 0.076) than in *H. teretifolia* (H_o average = 0.14; H_e average = 0.18; Figure 2). A positive relationship between F_{ST} and geographic distance was revealed among populations within both species, with a greater increase in F_{ST} over distance found in *H. sericea* compared with *H. teretifolia* (Figure 2a). We found a deficiency of heterozygotes in most populations of both species (observed heterozygosity was generally lower than expected under Hardy–Weinberg equilibrium), but six populations of *H. sericea* and three populations of *H. teretifolia* presented an excess of heterozygotes, resulting in negative F_{IS} estimates (Figure 2b).

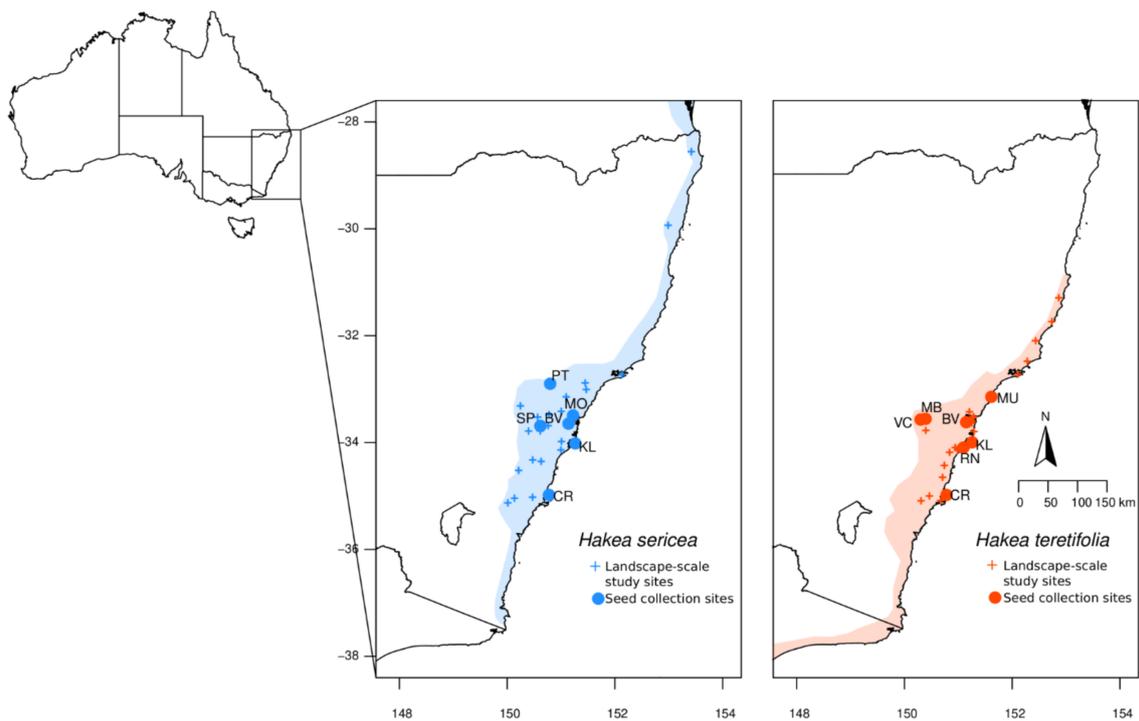


Figure 1. Maps showing distribution range of study species (shading), sites sampled for landscape-scale evaluation of genetic diversity (+), and sites sampled for progeny analysis (● and site code).

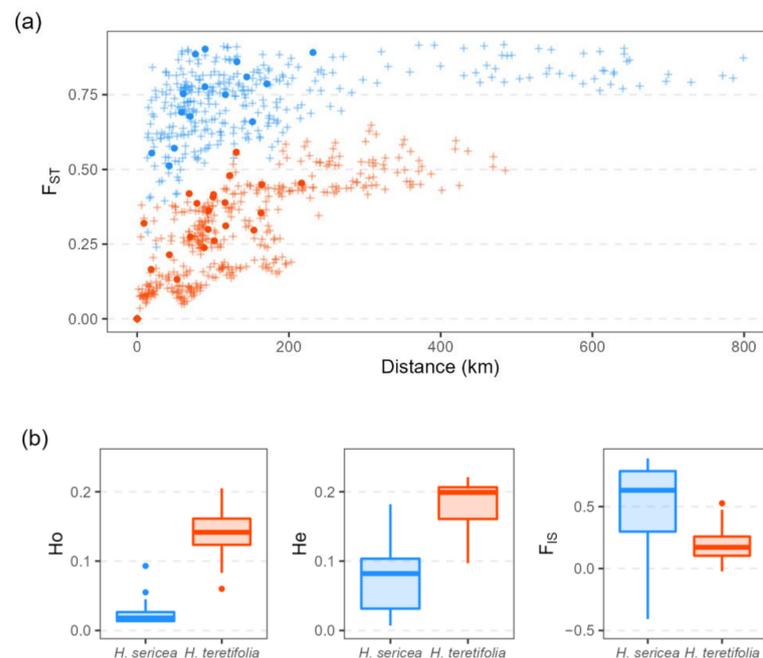


Figure 2. (a) Pairwise F_{ST} vs. geographic distance of *H. sericea* (blue) and *H. teretifolia* (orange) among populations sampled for landscape-scale evaluation of genetic diversity (+), and sites sampled for progeny analysis (●); (b) observed heterozygosity (H_o), expected heterozygosity (H_e), and F_{IS} estimates for populations sampled for landscape-scale evaluation of genetic diversity.

2.1.2. Mothers and Seedlings

A significantly higher overall germination success (proportion of total seeds germinated at the end of the experiment) was observed for *H. sericea* than *H. teretifolia* (94% vs. 52%, $p = 1.01 \times 10^{-43}$). The processed SNP data from seedling and mother samples

comprised 19,287 SNPs for *H. sericea* (371 total individuals across 6 sites), and 35,899 SNPs for *H. teretifolia* (376 total individuals across 7 sites). Estimates of H_o , H_e , and F_{IS} of the seed mothers mostly fell within the range of population genomics estimates from the broader geographic sampling of wild plants (Figure 2b and Table S3), though some sites were consistent outliers (e.g., at *H. sericea* sites BV, MO in the Hawkesbury River valley, H_o and H_e were higher than at other sites).

2.2. Mating Systems

2.2.1. Selfing Rate

The filtered data sets used for the estimation of selfing rates consisted of 5177 SNPs for *H. sericea* and 6990 SNPs for *H. teretifolia* following the exclusion of missing data. Methods for calculating selfing rate estimates (\hat{s}) using the MME and CNN approaches (hereafter \hat{s}_M and \hat{s}_C , respectively), including validations and comparisons, are described in detail as Supporting Information. *H. sericea* site PT gave anomalous values (Figure S1) associated with high error (particularly in \hat{s}_C ; Figure S2), possibly due to very low genetic variation at this site (Figure S3), and was thus excluded from further analyses of the mating system.

Estimated selfing rates were significantly higher on average in *H. sericea* compared with *H. teretifolia* using both MME and CNN approaches (\hat{s}_M : *H. sericea* 0.55, *H. teretifolia* 0.38, $p = 1.11 \times 10^{-8}$; \hat{s}_C : *H. sericea* 0.57, *H. teretifolia* 0.38, $p = 2.57 \times 10^{-12}$; Figure 3a). Within each species, estimates of selfing rates varied among and within sites (Figure 3b); in *H. sericea*, \hat{s}_M and \hat{s}_C were on average lower in progeny from sites CR and MO compared with other sites, and in *H. teretifolia* \hat{s}_M and \hat{s}_C , estimates were lower at sites MU and BV, and higher at MB. The range in estimates within sites and maternal lines also varied, in some cases as much as the range among sites (Figure 4).

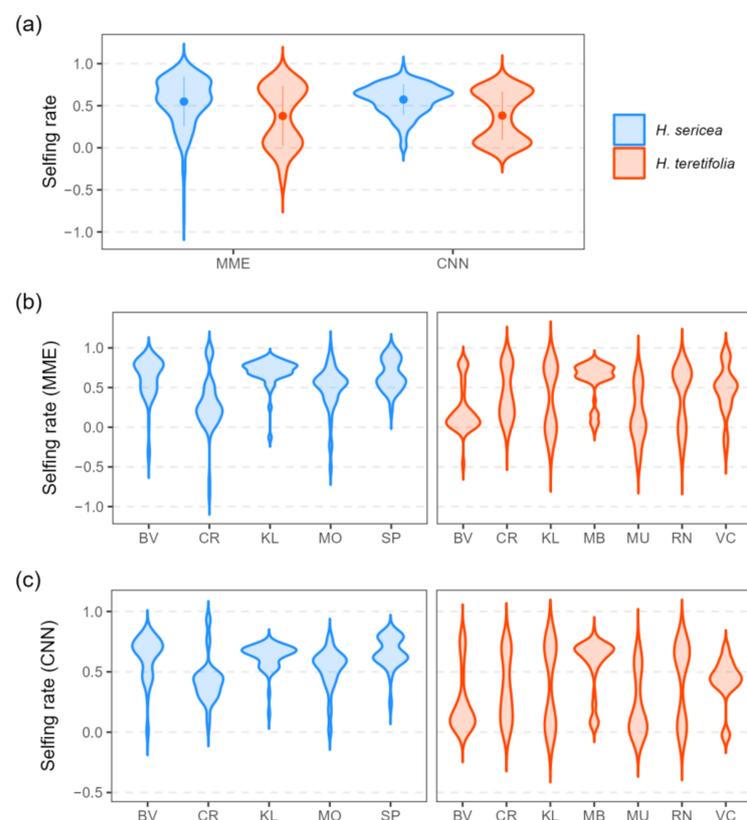


Figure 3. Violin plots (Gaussian kernel density estimates) of individual progeny selfing rates inferred using MME and CNN approaches for *H. sericea* (blue) and *H. teretifolia* (orange); per species (a), and per site (b,c).

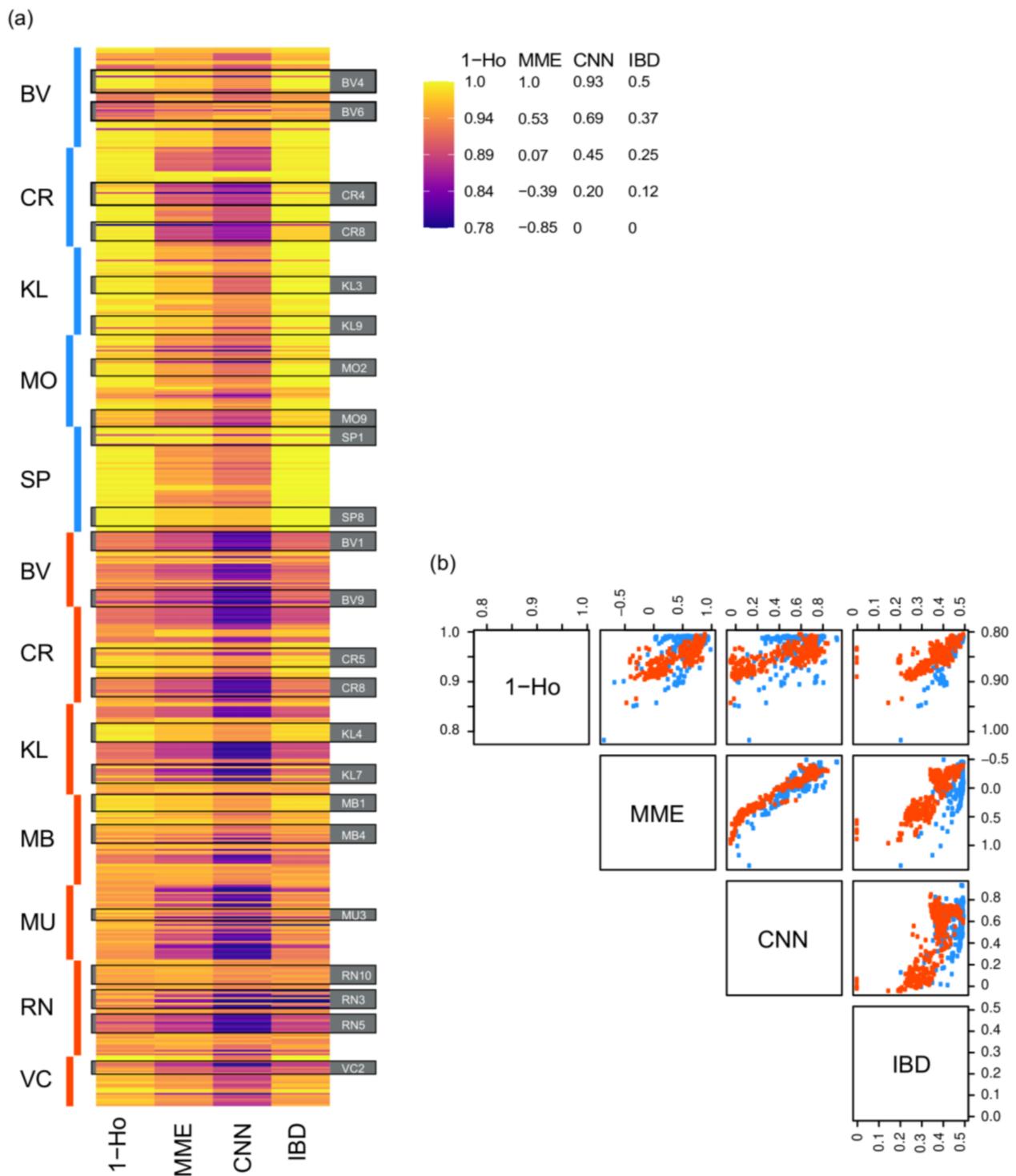


Figure 4. Comparisons among individual progeny of proportion of homozygous loci (1-Ho) and mating system estimates (MME: \hat{s}_M , CNN: \hat{s}_C , IBD: mother-progeny kinship coefficient, i.e., identity by descent). **(a)** Heatmap of scaled values side-by-side for each individual (horizontal stripe) grouped by species, site (vertical bar and 2-letter code), and maternal line (example maternal lines from each site are outlined and labeled with black boxes); **(b)** scatterplot matrix showing pairwise comparison of each set of individual values.

The two methods for inferring the mating system developed in this study were robust under the validation conditions used (see Supporting Information) and yielded largely concordant results (Figures 3 and 4). The means and ranges of \hat{s}_M and \hat{s}_C were similar among species and sites, but the shapes of the distributions differed. Distributions of selfing rate estimates within species and within sites were generally bimodal (tending to cluster around two modes), but the sizes and locations of the peaks varied among populations (Figure 3). At the species level, distributions of selfing rates had two distinct modes; in *H. sericea*, estimates were more frequently inferred to cluster around the higher selfing rate, whereas in *H. teretifolia*, values were more evenly distributed between a higher and a lower peak in selfing rate (Figure 3a). Similarly, clear bimodality was observed in estimates within most *H. teretifolia* sites, with relative sizes of peaks varying between sites (e.g., higher selfing rates more frequently inferred in site MB, and lower rates more frequently inferred in BV); however, peaks in distributions of estimates were less distinct within *H. sericea* sites (Figure 3b,c).

2.2.2. Relatedness

Relatedness estimates were consistent with theoretically inferred values (i.e., ≥ 0.25 between parents and progeny, and ≥ 0.125 between siblings with unknown paternity), and were significantly higher in *H. sericea* than in *H. teretifolia* (Figure S4). Pairwise kinship coefficients between mothers and progeny showed a positive relationship with selfing rate (Figure 4).

2.2.3. Patterns in Heterozygosity

Heterozygosity in the seed collection sites was lower in *H. sericea* relative to *H. teretifolia* (Table S3). There was broad concordance between the observed levels of heterozygosity and inferred mating systems; seedlings with relatively high proportions of heterozygous loci had lower selfing rate estimates (\hat{s}_M and \hat{s}_C), i.e., were inferred to be more likely to be outcrossed (Figure 4). Higher selfing was related to the loss of heterozygosity in seedlings relative to mothers, consistent with theoretical expectations (Figure S5); the difference in heterozygosity was negatively correlated with \hat{s}_M and \hat{s}_C in *H. teretifolia*, but in *H. sericea*, no significant correlation was found.

2.3. Effect of Sampling Strategy on Genetic Diversity in Subsamples of Seedlings

While sampling targets were the same among sites and between species, variation in germination success resulted in unequal numbers of seedlings being available for re-sampling. Germination success did not differ significantly among sites in *H. sericea*, but showed significant variation among *H. teretifolia* sites (e.g., CR 86%, MU and VC 18% and 20%, respectively; Table S2). Thus, there were too few *H. teretifolia* seedlings from sites MU and VC to meet resampling targets, and results for *H. teretifolia* are based on five sites (BV, CR, KL, MB, RN). Data were analysed from at least 9 maternal lines per site with at least 3 offspring each, and 3 maternal lines per site comprised at least 10 offspring (except *H. teretifolia* site BV, with only 2 maternal lines of at least 10 offspring; Table S1).

Within-site seedling subsamples which included more maternal lines had higher genetic diversity (within-subsample average distance) than those which included fewer maternal lines, in both species (Figure 5). At each site, targeting eight mothers resulted in the highest average genetic distances among offspring, whereas targeting only one mother resulted in the lowest (with the exception of *H. sericea* sites PT and SP; Figure 5).

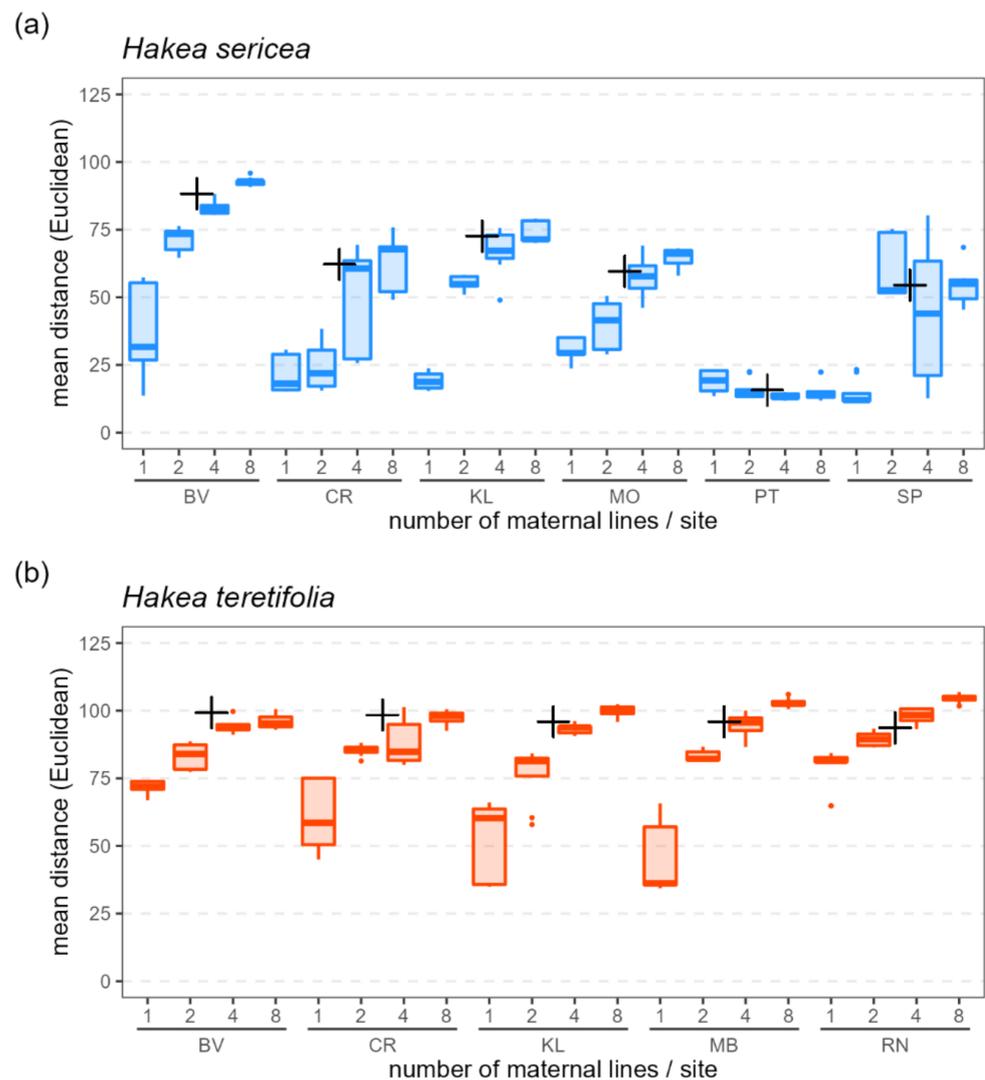


Figure 5. Genetic diversity (as average of pairwise Euclidean distances) among replicate subsamples of 8 seedlings per site for (a) *H. sericea* and (b) *H. teretifolia* according to 4 different sampling strategies (8 seedlings from 1 mother, 4 seedlings from 2 mothers, 2 seedlings from 4 mothers, 1 seedling from each of 8 mothers). Plotted for each site are averages among replicate sets of 8 seedlings randomly resampled 10 times per sampling scheme (boxplots), and average among the mother plants (+).

In both species, among-site seedling subsamples captured higher proportions of alleles when more sites were sampled (Figure 6). The maximum allele capture proportion increased as the minimum allele frequency threshold increased (i.e., as only increasingly common alleles were considered; Figure S6). For common alleles (those found on a landscape scale at $\geq 5\%$ frequency), the proportion captured was lower, and the range in proportion captured was higher, among *H. sericea* resampling replicates compared with *H. teretifolia* (this pattern remained consistent even when the two most geographically distant *H. sericea* populations were removed; data not shown). In *H. sericea*, allele capture proportion ranged from 47% (2 sites) to 91% (6 sites). In *H. teretifolia*, allele capture proportion ranged from 76% (2 sites) to 96% (5 sites). No resampling replicate captured 100% of the common alleles in either species. Capturing 90% of the common alleles required at least 3 sites in *H. teretifolia* and at least 6 sites in *H. sericea*; capturing 80% of the common alleles required at least 2 sites in *H. teretifolia* and at least 4 sites in *H. sericea* (Figure 6).

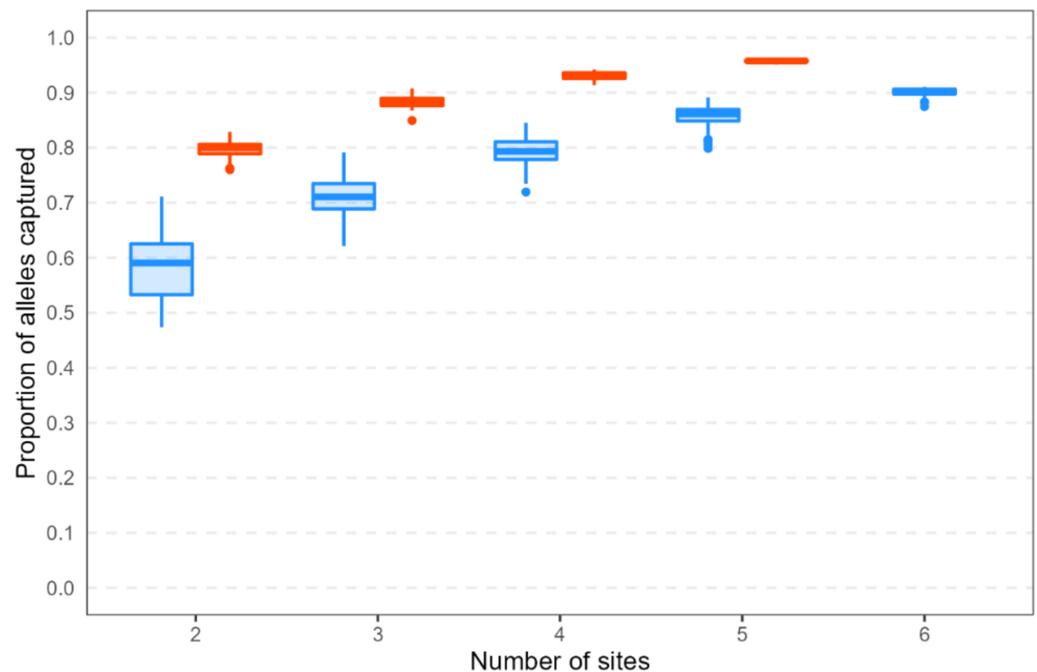


Figure 6. Proportions of alleles captured in seedling resampling replicates, out of all alleles found at $\geq 5\%$ frequency in adult plants (i.e., including landscape-scale sampling sites from which no progeny were genotyped). Shown are results from 100 resampling replicates for 8 seedlings per site representing 8 maternal lines per site across increasing numbers of collection sites.

3. Discussion

Empirical studies contribute towards understanding how genetic diversity is shaped within and among wild populations, and inform the development of practical, broadly applicable collecting guidelines to enhance the value of germplasm collections. Here, we gathered and analysed genomic data from two congeneric, seemingly functionally similar species, and developed tools enabling insight into mating systems using data from progeny analysis. This allowed comparisons exploring the extent to which population genetic patterns and reproductive behaviours are similar between related species (and among populations within species), as well as interpretation of the representative potential of seed collections and how to optimise these collections to best capture genetic diversity.

3.1. Contrasting Patterns of Genetic Diversity and Mating System in *H. sericea* vs. *H. teretifolia*

A clear contrast was revealed in the distribution of genetic diversity in the two species: low genetic diversity within populations (low heterozygosity, high F_{IS}) and high differentiation among populations (high F_{ST}) in *H. sericea* relative to *H. teretifolia*. Mating systems inferred for the two species from progeny analysis were consistent with the contrasting patterns in genetic diversity: while both species were capable of self-fertilisation and cross-fertilisation, prevalent selfing (and/or biparental inbreeding, discussed further below) was found in *H. sericea* (high selfing rate, high relatedness) relative to *H. teretifolia*. This relationship between patterns of genetic diversity and mating systems is consistent with expectations [8]. The strong contrast between the two species would not have been apparent without genetic data; *H. sericea* and *H. teretifolia* are similar in floral morphology and life history, and there are no prior reports on their mating systems (though variation in mating system among other species of *Hakea* has been recorded; [26,27]). While these contrasting patterns were unexpected given the outward similarities between *H. sericea* and *H. teretifolia*, they add to a growing body of findings demonstrating sometimes markedly contrasting patterns between species that are taxonomically related and otherwise functionally similar ([28,29]).

3.2. How Feasible Is It to Predict Mating System?

This study directly estimated selfing rates from progeny array datasets consisting of thousands of SNPs using novel methods (a reimplementation of a likelihood approach, and a Convolutional Neural Network). Results from the two approaches were concordant with one another under simulated scenarios, and with the empirical data, and were consistent with patterns of heterozygosity and estimated kinship coefficients, and we are optimistic that one or both of these approaches can extend to broader applications in the future.

The bimodality of selfing rate estimates is an intriguing pattern; one possible interpretation is that each peak corresponds to either outcrossed (lower \hat{s}_M and \hat{s}_C) or inbred (higher \hat{s}_M and \hat{s}_C) progeny. Anomalous \hat{s}_M and \hat{s}_C values from *H. sericea* site PT can be attributed to low heterozygosity together with exceptional genotypic uniformity leading to insufficient allele frequency variation for inference of selfing under expected rates of genotyping error. Possible explanations for this lack of diversity include multiple generations of inbreeding following a severe bottleneck (e.g., a single founder), and/or asexual reproduction (though we are unaware of prior reports of apomixis in *Hakea*); essentially, every plant sampled at this site belonged to a single isogenic (genetically identical) line. It is encouraging that error values associated with selfing estimates (particularly \hat{s}_C) suggested that inference for PT might be poorer than other populations.

While low selfing rate estimates indicate outcrossed seedlings (resulting from random mating within the population, e.g., in *H. sericea* site CR and *H. teretifolia* site BV), the specific process leading to high estimates can be harder to infer. When population heterozygosity levels are low and cross-fertilization occurs between close relatives, estimates of selfing rate can skew upward because biparentally inbred individuals can appear to be selfed [30–32]. Thus, while high \hat{s}_M and \hat{s}_C (such as in *H. sericea* generally and *H. teretifolia* site MB) are indicative of inbreeding, making the finer distinction between selfing and biparental inbreeding is difficult. We note, however, that this distinction is unlikely to impact seed sampling strategies for capturing genetic diversity at a site.

Consistent with theoretical expectations, species-level patterns of population genomic diversity (low heterozygosity, high F_{IS} , and high F_{ST} in *H. sericea* relative to *H. teretifolia*) reflected a predominantly selfing and/or biparentally inbreeding mating system in *H. sericea* and a mixed mating system more evenly balanced between selfing and outcrossing in *H. teretifolia*. This suggests that, at a species level, the predominant mating system could be predictable from patterns of genetic diversity such as are increasingly commonly characterized in plant conservation and restoration studies (e.g., [22,33]). However, while patterns in F_{IS} (i.e., differences between observed and expected heterozygosity under Hardy–Weinberg equilibrium) were consistent with inferred mating system variation in this study, other factors (e.g., purifying selection) may also contribute. We note the sensitivity of F_{IS} to small sample sizes such as used in this study [34], and potentially also to genotyping error [35], especially when genetic variation is low (such as at site PT).

At the population level, patterns of genetic diversity did not always show expected relationships with the mating system. For example, at *H. sericea* site BV, heterozygosity and selfing rates were both relatively high, and F_{IS} was relatively low, compared to other populations of this species. Population demographic history and environmental variation (both spatial and temporal; [36]) are likely explanations for these patterns. While detailed exploration of such processes was beyond the scope of this study, we observed that the mating system varied widely among, and sometimes within, families, suggesting that any individual of either study species may readily produce both selfed and outcrossed seeds depending on intrinsic or extrinsic factors yet to be characterised. It is intriguing that inland populations of *H. sericea* (PT, SP) were presumably more highly inbred than coastal sites, but a similar pattern was not observed in *H. teretifolia*; additional study is required to test the influence of habitat on selfing rate. Our findings provide a foundation for further work using *H. sericea* and *H. teretifolia* as model systems to investigate the processes determining selfing vs. outcrossing in mixed-mating species. Future studies

should focus on experimentation using increased within-population sampling for a subset of populations replicating particular environmental variables over time and space.

3.3. Genetic Diversity Captured in Seeds Reflects the Mating System of the Species

The within-site average pairwise genetic distance in subsamples of seedlings of both *H. sericea* and *H. teretifolia* increased as more maternal lines were sampled, and approached that of the wild mother plants when at least four maternal lines were sampled (Figure 5). We note, however, that the two *H. sericea* sites with the lowest observed heterozygosity did not follow this trend. Seedling subsamples from site PT lacked diversity regardless of the number of maternal lines represented, and seedling subsamples from site SP comprised greater genetic diversity when drawn from two or four maternal lines than eight. This effect is likely attributable to the mating system: repeated generations of selfing together with the fixation of alleles (e.g., via founder effects and/or selection) may result in a maternal line becoming so homozygous as to be essentially isogenic (genetically identical). Representing within-site genetic diversity is then a matter of distributing sampling effort to evenly capture all such lines present, noting that samples from multiple mother plants within the same isogenic line would be redundant for the purposes of capturing genetic diversity. Individuals from site PT comprised a single isogenic line (all were essentially genetically identical, and thus sampling from more mothers did not capture additional diversity in the seedlings), whereas site SP comprised four such lines (Figure S7).

Among sites, alleles found in the landscape-scale study of genetic diversity were better represented in resampled seedlings from the subset of seed collection sites in *H. teretifolia* than they were in *H. sericea*. In both species, the allele capture proportion increased as more sites were sampled, but the increase per site was greater in *H. sericea*. This reflects the distribution of genetic diversity within the two species, and is consistent with long-held expectations of the effects of mating system on population structure [8,11]. For regional-scale genetic diversity, the benchmark of capturing all common alleles (>5% frequency) with 95% probability [10] was not met by the subset of seed sampling sites in this study (five in *H. teretifolia*, six in *H. sericea*). Thus, our findings reinforce those of Neel and Cummings [13] in reporting that seed from five sites is insufficient to represent species-level genetic diversity, regardless of the mating system. Further, we showed empirically that it was necessary to sample more populations of a selfing species compared with a mixed mating species to maximise species-level genetic diversity in seed collections. More specifically, we found that twice the number of sites was required in the selfing species to reach the same proportion of allele capture as the mixed-mating species. Thus, we recommend that germplasm collections aiming to maximise species-level genetic diversity should double the number of populations sampled of a selfing species relative to an outcrossing or mixed-mating species.

The implications of highly structured population genetic diversity for making representative seed collections have been long understood, e.g., Brown and Marshall [37] suggested that among-site sampling should be prioritised in selfing species compared with outcrossing species, even at the cost of within-site sampling. However, this recommendation is explicitly restated in only two [16,38] out of six [3,17,39,40] recently published seed sourcing guidelines for germplasm conservation. In practice, there may be limited capacity to take mating systems into account (especially since mating systems are generally unknown for most species), which may explain the inconsistent dissemination of Brown and Marshall's [37] advice. Ours is one of two recent studies explicitly testing and confirming this advice using empirical data: we inferred otherwise cryptic mating systems from SNP data, whereas Kallow et al. [19] had prior knowledge of the mating system. Given the concordance between patterns of genetic diversity and mating systems in both studies, and increasing trends towards gathering landscape-scale genomic data for plant conservation and restoration, the feasibility of predicting mating systems in data-deficient species should improve in the near future. Improved characterisation of mating systems should foster improved consideration of them as part of sourcing genetically representative germplasm collections.

3.4. How Does Mating System Modulate the Benefits of Genetic Diversity in Restored Populations?

We recommend that the causes and consequences of inbreeding (whether selfing or biparental inbreeding) be considered on a case-by-case basis. More specifically, we emphasize the distinction between a habitual inbreeder which has no significant loss of heterozygosity in selfed progeny and has presumably purged its genetic load at the population level (e.g., *H. sericea*), and a species with a more balanced reproductive strategy which loses heterozygosity when selfed and might suffer from the detrimental effects of excessive inbreeding (e.g., *H. teretifolia*).

In *H. sericea*, inbreeding depression might not pose much threat to population persistence, and populations might be more likely to establish from a smaller starting effective population size (Baker's Law [41]; the success of *H. sericea* as an invasive species is consistent with this). Outbreeding depression [42] might warrant strategies to mitigate (e.g., local provenancing), on the other hand, maximising within-site diversity might secure greater resilience to change. While the long-term evolutionary consequences of habitual selfing as a reproductive strategy are likely to be negative [43], evolutionary timescales are usually beyond the scope of consideration when restoring populations from seed.

In contrast to the habitually selfing *H. sericea*, *H. teretifolia* might be more vulnerable to inbreeding depression; low germination rates at two *H. teretifolia* sites in this study could be a sign of this. We thus expect *H. teretifolia* to be an example of a mixed-mating species which will be negatively impacted by inbreeding, where it would be advisable to minimise relatedness when sourcing this species' seeds for restoration and seed-banking practices. Furthermore, maintaining pollinator availability in restored populations of the mixed-mating *Hakea* species might be a priority to ensure their longer-term success [44], but could be considered less relevant in populations of selfing species such as *H. sericea*.

Our study is the first to our knowledge to use genome-wide marker data to empirically investigate strategies for capturing genetic diversity in seeds from a preferentially selfing species compared with its mixed-mating congener. Results confirm longstanding expectations by demonstrating that contrasting patterns of genetic diversity in a selfing compared with a mixed-mating species warrant contrasting collecting strategies (in terms of sampling more vs. fewer populations). Improved characterisation and consideration of mating systems in the design of germplasm collections will directly enhance the long-term viability of populations restored from such collections. Further studies focusing on the relationship between mating systems and genetic diversity in a germplasm conservation context should evaluate a varied range of species to ensure future seed-sourcing guidelines are broadly applicable.

4. Materials and Methods

4.1. Study Species

Hakea sericea Schrad. & J.C.Wendl and *Hakea teretifolia* (Salisb.) Britten (Proteaceae) are common shrubs of temperate woodlands and heath, endemic to southeastern Australia. These two species represent two clades in *Hakea* which radiated independently in southeastern Australia approximately 30 (± 10) million years ago [45]. However, they share many morphological and functional traits, such as terete, needle-like leaves; small, white, zygomorphic flowers in axillary inflorescences; and woody, serotinous, two-seeded follicles (that of *H. sericea* is globose, 2–4 times thicker than the tapered follicle of *H. teretifolia*). Both species are serotinous, with a long-lived canopy seedbank and fire-enhanced dispersal, but are sensitive to high-intensity fires [46]. They can be found growing in sympatry on sandstone substrates in sclerophyll forests and heath where their distributions overlap (Figure 1). *Hakea sericea* is invasive in South Africa, New Zealand, and Europe.

4.2. Field Sampling

To characterise and compare landscape-scale patterns in population genomics within and between *H. sericea* and *H. teretifolia*, wild plants were sampled for genotyping across

the distribution of each species in New South Wales, Australia (26 sites for *H. sericea* and 28 for *H. teretifolia*; Figure 1, Table S1a) following the restore and renew methodology for field sampling, DNA preservation, and genotyping (six individuals per site; [22]).

To explore mating system variation and the extent to which seed collections captured genetic diversity, we subsequently collected seeds and mother plant tissue samples from multiple sites for each species (six *H. sericea*: BV, CR, KL, MO, PT, SP; and seven *H. teretifolia*: BV, CR, KL, MB, MU, RN, VC; Figure 1, Table S1b). Sites sampled included coastal and inland populations representing the full elevational range of each species within the Sydney Basin Bioregion (Table S1b). At each site, leaf tissue was collected from 10 plants, each at least 10 m apart to minimise the sampling of first-degree relatives, and processed as described above. Multiple fruits were collected from each plant and stored in paper bags until seed extraction. Fruits were dried at 30 degrees C to open the follicles; extracted seeds were stored in paper envelopes under climate-controlled conditions until sowing.

4.3. Seed Germination and Seedling Sampling

Seed germination and seedling propagation were conducted at the Australian Plant-Bank. Seeds were surface sterilised by rinsing in 2% bleach for 3 min, then rinsing in autoclaved, deionised water, and immediately sown on 1% agar in petri dishes and incubated at a 12h/12h light/dark cycle. Seedlings were potted in forestry tubes, and two-to-three adult leaves were harvested from each seedling for DNA extractions and genotyping. Individuals were selected for genotyping to allow evaluation of the effect of sampling strategy on seedling genetic diversity relative to mothers: 10 maternal lines per site per species, with 7 represented by the mother and 3 progeny, and 3 represented by the mother and 10 progeny. This sampling was designed to enable replicate subsets of seedling data to represent a range of hypothetical seed sampling strategies, from few seeds/many mothers to many seeds/few mothers, within and among sites. When the number of available seedling tissue samples was insufficient to meet the target, all available samples were submitted (Table S2).

4.4. Genomic Data Collection and Processing

Lyophilised leaf tissues were randomized in 96-well tube racks and submitted to Diversity Arrays Technology Pty Ltd. (DART) for DNA extraction and genotyping using the medium density DArTseq option. The DART pipeline has successfully informed many studies across a wide range of model and non-model organisms (e.g., [47,48]) and is a cost- and time-efficient solution for generating genomic-scale data. Genotype data were provided as Single Nucleotide Polymorphism (SNP) matrices assembled through DART's proprietary analytical pipeline supplied with quality informing statistics for each bi-allelic marker and sample. In total, 1156 plants were genotyped using DArTseq.

Data were processed and downstream analyses were carried out using R. Prior to using the SNP data for downstream analysis, matrices were filtered based on several quality criteria using an in-house R package (see [22]). In summary, SNPs with a reproducibility score below 98% and missing genotype calls for more than 10% of samples were removed; to minimize the effect of linkage, only one SNP was maintained from markers with more than one SNP, and samples with more than 40% missing data were excluded from further analysis.

4.5. Estimates of Genetic Diversity and Relatedness

Genetic diversity measures were calculated from the filtered SNP data using the R packages *diveRsity* [49] (observed and expected heterozygosity, F_{IS}) and *SNPrelate* [50] (population-pairwise F_{ST}).

Pairwise kinship coefficients were estimated as probabilities that loci were identical by descent (IBD in the R package *SNPrelate*; [50]), using allele frequencies for each population estimated from the wild-sampled mother plants of that population (custom script). Kinship coefficients between species were compared using Mann–Whitney U tests.

4.6. Inference of Selfing Rates

We aimed to investigate mating systems in the context of seed collections for restoration and conservation, specifically asking how many of the progeny collected from 10 known mothers per site were produced via selfing vs. outcrossing. Several software tools exist for investigating mating systems, including MLTR ([25]; also see [51]) and COLONY ([52]). However, since MLTR was designed for microsatellite datasets (usually consisting of dozens to hundreds of markers each with multiple alleles), it was difficult to take advantage of our full SNP datasets (consisting of thousands of biallelic markers) using MLTR. The package COLONY usually relies on near-comprehensive population sampling, whereas we sampled individuals representing only a small fraction of the population at each site. We therefore developed tools for estimating the selfing rate (s) for each progeny that were suited to the available data. First, we implemented a single-locus method of moments estimator of selfing (MME) that was presented by Ritland [25]. We tested this implementation by generating progeny datasets that were simulated with different levels of selfing, but that reflected other features of the data, including the observed maternal genotypes and population allele frequencies, and estimated rates of genotyping error [53,54]. We checked that values inferred using the estimator were similar to those used in the corresponding simulations (see Supplementary Methods). Second, we performed more simulations of progeny data under different levels of selfing, and trained Convolutional Neural Network (CNN) models to predict rates of selfing. We then used the models to predict selfing rates for empirical progeny genotype data. Our goal was to develop an approach for mating system estimation that could be applied to many loci, and that, in the future, might be extended to data with different properties (e.g., with ordered markers) or to inferring different properties of mating systems (e.g., progeny produced by apomixis). Recent reports have shown that CNNs can be highly effective and flexible when applied to suitable genetic data and problems (e.g., [55–57]). Methods for calculating selfing rate estimates (\hat{s}) using the MME and CNN approaches (\hat{s}_M and \hat{s}_C , respectively), including validations and comparisons, are described in detail as Supporting Information (Supplement S1).

Statistical significance of differences in \hat{s}_M and \hat{s}_C between species was calculated using Mann–Whitney U tests. For each seedling, we then compared estimated selfing rates to mother–seedling difference in heterozygosity (proportion of heterozygous loci in seedling–proportion of heterozygous loci in mother) using Spearman’s correlation coefficient (ρ).

4.7. Exploration of Seed Sampling Strategies

We used progeny array data to evaluate the range in genetic diversity captured by subsamples of seedlings according to different hypothetical sampling strategies (i.e., distributing sampling effort among fewer vs. more maternal lines and sites). We resampled replicate subsets of seedlings for each species by randomly drawing seedlings among maternal lines and sites according to the sampling schemes described below (custom scripts).

The within-site sampling schemes drew ten replicate subsamples of eight seedlings: one seedling from each of eight mother plants, or two seedlings from each of four mother plants, or four seedlings from each of two mother plants, or eight seedlings from a single mother plant. For each replicate subsample, pairwise Euclidean distance matrices were calculated and the mean of each matrix taken to express the genetic diversity captured per subsample per site. Euclidean distances were chosen to measure genetic diversity for their simplicity, lack of assumptions, and demonstrated concordance with other more complex metrics [58]; small within-site sample sizes together with unknown rates of genotyping error precluded the use of other metrics (such as counts of alleles captured).

Among-site sampling schemes drew 100 replicate seedling subsamples from among all seed collection sites per species. We compared the effect of sampling fewer vs. more sites with a scheme drawing one seedling from each of eight maternal lines per site, across increasing numbers of seed collection sites. For among-site subsamples, we evaluated allele capture by calculating whether both alleles at polymorphic loci in adult wild-sampled

individuals were also present in each seedling subsample. We considered alleles occurring at $\geq 0.5\%$, 1%, 2%, and 5% frequency in mother plants (seed collection sites only), and all wild plants (landscape-scale study sites as well as seed collection sites). Alleles rarer than 0.5% were not considered due to the possibility of genotyping error in this frequency range.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12030522/s1>: Study Data; Supplementary Methods; Table S1: Sample collection sites and details; Table S2: Seed (progeny array) details and germination results; Table S3: Estimates of genetic diversity for seed sampling sites; Figure S1: Individual progeny selfing rates inferred per site (including site PT); Figure S2: Measures of error in selfing rate estimates; Figure S3: Genetic distance matrices among *H. sericea* individuals from sites PT and SP; Figure S4: Progeny array pairwise kinship coefficient estimates by known relationship; Figure S5: Loss of heterozygosity in progeny by selfing rate; Figure S6: Maximum proportions of alleles captured by resampling seedlings.

Author Contributions: Conceptualization, M.R. and M.M.v.d.M.; data curation, P.L.-I., K.K. and M.M.v.d.M.; formal analysis, P.L.-I., J.G.B., M.O. and M.M.v.d.M.; funding acquisition, M.M.v.d.M.; investigation, P.L.-I., J.G.B., K.K. and M.M.v.d.M.; methodology, P.L.-I., J.G.B. and M.M.v.d.M.; project administration, P.L.-I. and M.M.v.d.M.; resources, M.R. and M.M.v.d.M.; software, J.G.B.; supervision, M.R. and M.M.v.d.M.; visualization, P.L.-I. and M.O.; writing—original draft, P.L.-I., J.G.B. and M.M.v.d.M.; writing—review & editing, P.L.-I., J.G.B., M.R., K.K., M.O. and M.M.v.d.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the NSW Environmental Trust (grant 2016RD 0084).

Data Availability Statement: The data presented in this study are available as Supplementary Materials (Study Data). The code used for the inference of mating systems is accessible via a Zenodo repository: <https://doi.org/10.5281/zenodo.7547080>.

Acknowledgments: We acknowledge the traditional custodians of the land on which this research was conducted. We thank all volunteers and colleagues who assisted, particularly Joel Cohen, Jamilly Machado da Silva, Richard Dimon, Hannah McPherson, Lotte von Richter, Veronica Viler, and staff of the Australian PlantBank. We thank the staff of the New South Wales National Parks and Wildlife Service for access to sampling sites.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Broadhurst, L.M.; Lowe, A.; Coates, D.J.; Cunningham, S.A.; McDonald, M.; Vesk, P.A.; Yates, C. Seed Supply for Broad-scale Restoration: Maximizing Evolutionary Potential. *Evol. Appl.* **2008**, *1*, 587–597. [[CrossRef](#)]
2. Hoffmann, A.A.; Sgrò, C.M. Climate Change and Evolutionary Adaptation. *Nature* **2011**, *470*, 479–485. [[CrossRef](#)] [[PubMed](#)]
3. Harrison, P.A.; Breed, M.F.; Broadhurst, L.; Byrne, M.; Commander, L.; Coates, D.; Krauss, S.; van der Merwe, M.; Miller, M.; Rossetto, M. Florabank Guidelines Module 5: Seed Sourcing. In *Florabank Guidelines: Best Practice Guidelines for Native Seed Collection and Use*; Australian Government, Florabank Consortium: Canberra, Australia, 2021.
4. Young, A.G.; Pickup, M. Low S-Allele Numbers Limit Mate Availability, Reduce Seed Set and Skew Fitness in Small Populations of a Self-Incompatible Plant. *J. Appl. Ecol.* **2010**, *47*, 541–548. [[CrossRef](#)]
5. Weeks, A.R.; Sgrò, C.M.; Young, A.G.; Frankham, R.; Mitchell, N.J.; Miller, K.A.; Byrne, M.; Coates, D.J.; Eldridge, M.D.B.; Sunnucks, P.; et al. Assessing the Benefits and Risks of Translocations in Changing Environments: A Genetic Perspective. *Evol. Appl.* **2011**, *4*, 709–725. [[CrossRef](#)] [[PubMed](#)]
6. Gavin-Smyth, N.; Kramer, A.T.; Urbina-Casanova, R.; Vitt, P.; Fant, J.B. Genetic Rescue Reduces Mate Limitation in a Threatened, Clonal, and Self-incompatible Plant Species. *Restor. Ecol.* **2021**, *29*, e13458. [[CrossRef](#)]
7. Hamrick, J.L. Plant Population Genetics and Evolution. *Am. J. Bot.* **1982**, *69*, 1685–1693. [[CrossRef](#)]
8. Charlesworth, D. Effects of Inbreeding on the Genetic Diversity of Populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 1051–1070. [[CrossRef](#)]
9. Barrett, S.C.H.; Arunkumar, R.; Wright, S.I. The Demography and Population Genomics of Evolutionary Transitions to Self-Fertilization in Plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2014**, *369*, 20130344. [[CrossRef](#)]
10. Marshall, D.R.; Brown, A.H.D. Optimum Sampling Strategy in Genetic Conservation. In *Crop Genetic Resources for Today and Tomorrow*; Frankel, O.H., Hawkes, J.G., Eds.; Cambridge University Press: London, UK, 1975; pp. 53–80.
11. Schoen, D.J.; Brown, A.H. Intraspecific Variation in Population Gene Diversity and Effective Population Size Correlates with the Mating System in Plants. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4494–4497. [[CrossRef](#)]

12. Hoban, S.; Schlarbaum, S. Optimal Sampling of Seeds from Plant Populations for Ex-Situ Conservation of Genetic Biodiversity, Considering Realistic Population Structure. *Biol. Conserv.* **2014**, *177*, 90–99. [[CrossRef](#)]
13. Neel, M.C.; Cummings, M.P. Effectiveness of Conservation Targets in Capturing Genetic Diversity. *Conserv. Biol.* **2003**, *17*, 219–229. [[CrossRef](#)]
14. Brown, A.H.D.; Briggs, J.D. Sampling Strategies for Genetic Variation in Ex-Situ Collections of Endangered Plant Species. In *Genetics and Conservation of Rare Plants*; Falk, D.A., Holsinger, K.E., Eds.; Center for Conservation, Oxford University Press: New York, NY, USA, 1991; pp. 99–119.
15. *Ex Situ Plant Conservation: Supporting Species Survival in the Wild*; Guerrant, E.O.; Havens, K.; Maunder, M. (Eds.) Island Press: Washington, DC, USA, 2004.
16. Center for Plant Conservation. *CPC Best Plant Conservation Practices to Support Species Survival in the Wild*; Center for Plant Conservation: Escondido, CA, USA, 2019.
17. Bureau of Land Management. *Technical Protocol for the Collection, Study, and Conservation of Seeds from Native Plant Species for Seeds of Success*; Bureau of Land Management: Phoenix, AZ, USA, 2018.
18. Hoban, S.; Strand, A. Ex Situ Seed Collections Will Benefit from Considering Spatial Sampling Design and Species' Reproductive Biology. *Biol. Conserv.* **2015**, *187*, 182–191. [[CrossRef](#)]
19. Kallow, S.; Panis, B.; Vu, D.T.; Vu, T.D.; Paofa, J.; Mertens, A.; Swennen, R.; Janssens, S.B. Maximizing Genetic Representation in Seed Collections from Populations of Self and Cross-Pollinated Banana Wild Relatives. *BMC Plant Biol.* **2021**, *21*, 415. [[CrossRef](#)]
20. Allendorf, F.W.; Hohenlohe, P.A.; Luikart, G. Genomics and the Future of Conservation Genetics. *Nat. Rev. Genet.* **2010**, *11*, 697–709. [[CrossRef](#)] [[PubMed](#)]
21. Prober, S.; Byrne, M.; McLean, E.; Steane, D.; Potts, B.; Vaillancourt, R.; Stock, W. Climate-Adjusted Provenancing: A Strategy for Climate-Resilient Ecological Restoration. *Front. Ecol. Evol.* **2015**, *3*, 65. [[CrossRef](#)]
22. Rossetto, M.; Bragg, J.; Kilian, A.; McPherson, H.; van der Merwe, M.; Wilson, P.D. Restore and Renew: A Genomics-Era Framework for Species Provenance Delimitation: Species Provenance Delimitation Framework. *Restor. Ecol.* **2018**, *26*, 420.
23. DeWoody, J.A.; Harder, A.M.; Mathur, S.; Willoughby, J.R. The Long-Standing Significance of Genetic Diversity in Conservation. *Mol. Ecol.* **2021**, *30*, 4147–4154. [[CrossRef](#)]
24. Kardos, M.; Armstrong, E.E.; Fitzpatrick, S.W.; Hauser, S.; Hedrick, P.W.; Miller, J.M.; Tallmon, D.A.; Funk, W.C. The Crucial Role of Genome-Wide Genetic Variation in Conservation. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104642118. [[CrossRef](#)]
25. Ritland, K. Extensions of Models for the Estimation of Mating Systems Using N Independent Loci. *Heredity* **2002**, *88*, 221–228. [[CrossRef](#)]
26. Lamont, B.B.; Olesen, J.M.; Briffa, P.J. Seed Production, Pollinator Attractants and Breeding System in Relation to Fire Response—Are There Reproductive Syndromes among Co-Occurring Proteaceous Shrubs? *Aust. J. Bot.* **1998**, *46*, 377–385. [[CrossRef](#)]
27. Moodley, D.; Geerts, S.; Richardson, D.M.; Wilson, J.R.U. The Importance of Pollinators and Autonomous Self-Fertilisation in the Early Stages of Plant Invasions: Banksia and Hakea (Proteaceae) as Case Studies. *Plant Biol.* **2016**, *18*, 124–131. [[CrossRef](#)] [[PubMed](#)]
28. Rossetto, M.; Wilson, P.D.; Bragg, J.; Cohen, J.; Fahey, M.; Yap, J.-Y.S.; van der Merwe, M. Perceptions of Similarity Can Mislead Provenancing Strategies—An Example from Five Co-Distributed Acacia Species. *Diversity* **2020**, *12*, 306. [[CrossRef](#)]
29. Hoban, S.; Callicrate, T.; Clark, J.; Deans, S.; Dosmann, M.; Fant, J.; Gailing, O.; Havens, K.; Hipp, A.L.; Kadav, P.; et al. Taxonomic Similarity Does Not Predict Necessary Sample Size for Ex Situ Conservation: A Comparison among Five Genera. *Proc. Biol. Sci.* **2020**, *287*, 20200102. [[PubMed](#)]
30. Ennos, R.A.; Clegg, M.T. Effect of Population Substructuring on Estimates of Outcrossing Rate in Plant Populations. *Heredity* **1982**, *48*, 283–292. [[CrossRef](#)]
31. Schoen, D.J.; Clegg, M.T. Estimation of Mating System Parameters When Outcrossing Events Are Correlated. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 5258–5262. [[CrossRef](#)]
32. Ritland, K. The Effective Proportion of Self-Fertilization with Consanguineous Matings in Inbred Populations. *Genetics* **1984**, *106*, 139–152. [[CrossRef](#)]
33. Rossetto, M.; Yap, J.-Y.S.; Lemmon, J.; Bain, D.; Bragg, J.; Hogbin, P.; Gallagher, R.; Rutherford, S.; Summerell, B.; Wilson, T.C. A Conservation Genomics Workflow to Guide Practical Management Actions. *Glob. Ecol. Conserv.* **2021**, *26*, e01492. [[CrossRef](#)]
34. Aguirre-Liguori, J.A.; Luna-Sánchez, J.A.; Gasca-Pineda, J.; Eguiarte, L.E. Evaluation of the Minimum Sampling Design for Population Genomic and Microsatellite Studies: An Analysis Based on Wild Maize. *Front. Genet.* **2020**, *11*, 870. [[CrossRef](#)]
35. Chen, B.; Cole, J.W.; Grond-Ginsbach, C. Departure from Hardy Weinberg Equilibrium and Genotyping Error. *Front. Genet.* **2017**, *8*, 167. [[CrossRef](#)]
36. Whitehead, M.R.; Lanfear, R.; Mitchell, R.J.; Karron, J.D. Plant Mating Systems Often Vary Widely Among Populations. *Front. Ecol. Evol.* **2018**, *6*, 38. [[CrossRef](#)]
37. Brown, A.H.D.; Marshall, D.R. A Basic Sampling Strategy: Theory and Practice. In *Collecting Plant Genetic Diversity: Technical Guidelines*; Guarino, L., Rao, V.R., Reid, R., Eds.; CAB International: Wallingford, UK, 1995; pp. 75–91.
38. ENSCONET. *ENSCONET Seed Collecting Manual for Wild Species*; Royal Botanic Gardens, Kew, Universidad Politécnica de Madrid, Eds.; Royal Botanic Gardens: Kew, UK, 2009.

39. Gann, G.D.; McDonald, T.; Walder, B.; Aronson, J.; Nelson, C.R.; Jonson, J.; Hallett, J.G.; Eisenberg, C.; Guariguata, M.R.; Liu, J.; et al. International Principles and Standards for the Practice of Ecological Restoration. *Second Edition. Restor. Ecol.* **2019**, *27*, S1–S46. [[CrossRef](#)]
40. Food and Agriculture Organization of the United Nations. *Genebank Standards for Plant Genetic Resources for Food and Agriculture*; Food & Agriculture Organization: Rome, Italy, 2018; ISBN 9789251082621.
41. Baker, H.G. Self-Compatibility and Establishment After “Long-Distance” Dispersal. *Evolution* **1955**, *9*, 347–349.
42. Frankham, R.; Ballou, J.D.; Eldridge, M.D.B.; Lacy, R.C.; Ralls, K.; Dudash, M.R.; Fenster, C.B. Predicting the Probability of Outbreeding Depression. *Conserv. Biol.* **2011**, *25*, 465–475. [[CrossRef](#)] [[PubMed](#)]
43. Wright, S.I.; Kalisz, S.; Slotte, T. Evolutionary Consequences of Self-Fertilization in Plants. *Proc. Biol. Sci.* **2013**, *280*, 20130133. [[CrossRef](#)] [[PubMed](#)]
44. Fernandes, R.E.; Millar, M.A.; Coates, D.J.; Byrne, M.; Krauss, S.L.; Williams, M.; Hopper, S.D. Plant Mating System Dynamics in Restoration: A Comparison of Restoration and Remnant Populations of *Hakea Laurina* (Proteaceae). *Restor. Ecol.* **2021**, *30*, e13609. [[CrossRef](#)]
45. Cardillo, M.; Weston, P.H.; Reynolds, Z.K.M.; Olde, P.M.; Mast, A.R.; Lemmon, E.M.; Lemmon, A.R.; Bromham, L. The Phylogeny and Biogeography of *Hakea* (Proteaceae) Reveals the Role of Biome Shifts in a Continental Plant Radiation. *Evolution* **2017**, *71*, 1928–1943. [[CrossRef](#)]
46. Bradstock, R.A.; Gill, A.M.; Hastings, S.M.; Moore, P.H.R. Survival of Serotinous Seedbanks during Bushfires: Comparative Studies of *Hakea* Species from Southeastern Australia. *Austral. Ecol.* **1994**, *19*, 276–282. [[CrossRef](#)]
47. Van der Merwe, M.M.; Yap, J.-Y.S.; Wilson, P.D.; Murphy, H.T.; Ford, A. All Populations Matter: Conservation Genomics of Australia’s Iconic Purple Wattle, *Acacia Purpureopetala*. *Diversity* **2021**, *13*, 139. [[CrossRef](#)]
48. Rutherford, S.; van der Merwe, M.; Wilson, P.G.; Kooyman, R.M.; Rossetto, M. Managing the Risk of Genetic Swamping of a Rare and Restricted Tree. *Conserv. Genet.* **2019**, *20*, 1113–1131. [[CrossRef](#)]
49. Keenan, K.; McGinnity, P.; Cross, T.F.; Crozier, W.W.; Prodöhl, P.A. diveRsity: AnRpackage for the Estimation and Exploration of Population Genetics Parameters and Their Associated Errors. *Methods Ecol. Evol.* **2013**, *4*, 782–788. [[CrossRef](#)]
50. Zheng, X.; Levine, D.; Shen, J.; Gogarten, S.M.; Laurie, C.; Weir, B.S. A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* **2012**, *28*, 3326–3328. [[CrossRef](#)] [[PubMed](#)]
51. Koelling, V.A.; Monnahan, P.J.; Kelly, J.K. A Bayesian Method for the Joint Estimation of Outcrossing Rate and Inbreeding Depression. *Heredity* **2012**, *109*, 393–400. [[CrossRef](#)] [[PubMed](#)]
52. Jones, O.R.; Wang, J. COLONY: A Program for Parentage and Sibship Inference from Multilocus Genotype Data. *Mol. Ecol. Resour.* **2010**, *10*, 551–555. [[CrossRef](#)] [[PubMed](#)]
53. Bresadola, L.; Link, V.; Buerkle, C.A.; Lexer, C.; Wegmann, D. Estimating and Accounting for Genotyping Errors in RAD-Seq Experiments. *Mol. Ecol. Resour.* **2020**, *20*, 856–870. [[CrossRef](#)] [[PubMed](#)]
54. Luca, F.; Hudson, R.R.; Witonsky, D.B.; Di Rienzo, A. A Reduced Representation Approach to Population Genetic Analyses and Applications to Human Evolution. *Genome Res.* **2011**, *21*, 1087–1098. [[CrossRef](#)]
55. Fligel, L.; Brandvain, Y.; Schrider, D.R. The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. *Mol. Biol. Evol.* **2019**, *36*, 220–238. [[CrossRef](#)]
56. Fonseca, E.M.; Colli, G.R.; Werneck, F.P.; Carstens, B.C. Phylogeographic Model Selection Using Convolutional Neural Networks. *Mol. Ecol. Resour.* **2021**, *21*, 2661–2675. [[CrossRef](#)]
57. Perez, M.F.; Bonatelli, I.A.S.; Romeiro-Brito, M.; Franco, F.F.; Taylor, N.P.; Zappi, D.C.; Moraes, E.M. Coalescent-Based Species Delimitation Meets Deep Learning: Insights from a Highly Fragmented Cactus System. *Mol. Ecol. Resour.* **2022**, *22*, 1016–1028. [[CrossRef](#)]
58. Shirk, A.J.; Landguth, E.L.; Cushman, S.A. A Comparison of Individual-Based Genetic Distance Metrics for Landscape Genetics. *Mol. Ecol. Resour.* **2017**, *17*, 1308–1317. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.