

Floristic classifications and bioregionalizations are not predictors of intra-specific evolutionary patterns

Received: 29 August 2023

Accepted: 22 November 2024

Published online: 30 December 2024

 Check for updatesPatrick S. Fahey^{1,2,3}✉, Richard J. Dimon^{1,2}, Marlien M. van der Merwe^{1,2}, Jason G. Bragg^{2,4} & Maurizio Rossetto^{1,2}✉

The relationship between intra-specific and inter-specific patterns and processes over evolutionary time is key to ecological investigations. We examine this relationship taking an approach of focussing on the association between vegetation and floristic classifications, summaries of inter-specific processes, and intra-specific genetic structuring. Applying an innovative, multispecies, and standardised population genomic approach, we test the relationship between vegetation mapping schemes and structuring of genetic variation across a large, environmentally heterogenous region in eastern Australia. We show that intra-specific genetic variation shows limited correspondence to vegetation and floristic classifications and is better explained by distance between sampled populations and the location of biogeographical features which limit gene flow. Mapping schemes with contiguous mapping classes, particularly larger ones, were more predictive of genetic lineages, whether based on environmental factors or not, than geographically non-contiguous schemes. We conclude that vegetation and floristic classifications are not closely correlated with intra-specific genetic patterns, showing that intra-specific processes are not recapitulated by inter-specific floristic assembly processes. This study showcases the need to implement landscape level evolutionary patterns, based on species specific datasets, in restoration and conservation activities.

An ongoing area of inquiry in ecology and evolutionary biology is the relationship between vegetation assembly patterns and intra-specific evolutionary and adaptive processes, and whether they are responding to the same landscape level processes^{1,2}. Previous works have presented contrasting hypotheses in this regard, some suggesting that inter-specific floristic assembly processes should encapsulate intra-specific patterns due to similar drivers³, others that intra-specific genetic processes are independent from inter-specific assembly patterns⁴, or that floristic assembly patterns are highly influenced by

individual species-level processes such as local occurrence, dispersal, and historical events⁵.

Here we address this key ecological question from the standpoint of asking whether floristic assemblage classifications and bioregionalisations act as predictors of species-level gene flow and boundaries between genetically defined lineages, and therefore, putatively, evolutionary adaptive patterns? We also address the relevant practical implications, i.e., is there any correlation between management areas and genetic lineages across diverse plant taxa?

¹Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Queensland 4072, Australia. ²Research Centre for Ecosystem Resilience, Botanic Gardens of Sydney, Sydney, New South Wales 2000, Australia. ³Queensland Herbarium & Biodiversity Science, Department of the Environment, Tourism, Science & Innovation, Toowoong, Queensland 4066, Australia. ⁴School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2033, Australia. ✉e-mail: patrick.fahey@des.qld.gov.au; maurizio.rossetto@botanicgardens.nsw.gov.au

Vegetation type classifications, bioregionalisations, and management area designations, designed to distil complex ecosystems into forms that can be easily understood and communicated by scientists, managers, and landholders, play a crucial role in biodiversity management^{6–9}. However, a tension exists between these human classifications and how natural landscape-level processes function; political, economic and management systems are generally based upon hard borders while natural systems tend to operate as gradients over space and time with few clear boundaries^{10–12}.

These artificial landscape divisions are often reinforced with such regularity that we put weight on conserving their distinctiveness in and of itself, despite primarily reflecting a human view of the landscape either as political boundaries or as interpretations of assembly patterns^{12,13}. However, it is important to consider how strong the relationship is between our anthropocentric schemes and natural systems, as management of functioning ecosystems is crucial to ensure the sustainable future of natural environments and human flourishing as the effects of climate change and continued global disruption play out^{14,15}. Mismanagement can result in the systematic degradation of processes (such as maintenance of genetic diversity and gene flow) vital for the preservation and enhancement of the evolutionary potential and adaptability of restored and natural populations¹⁶. A previous study investigating relationships between seed transfer zones (STZ) and genetic structure concluded that while substantial genetic diversity is captured using these zones the genetic structuring were species specific and did not necessarily align with that of the STZs¹⁷.

Here we assess if vegetation classification and bioregionalisation schemes correspond to the natural landscape-genetic patterns estimated from genome wide molecular data obtained via standardised sampling¹⁸ for 50 widespread plant species co-occurring across a large, heterogeneous and highly biodiverse geographical area. The study system includes multiple climatic gradients and vegetational shifts covering a broad range of biomes from rainforests to semi-arid woodlands within the state of New South Wales (NSW), Australia (Fig. 1). A hierarchical vegetation and floristic community classification scheme^{6,19} and a nested bioregionalisation scheme²⁰, both used in management decision making within the study region, were chosen for investigation (Fig. 1b–d). Additionally, two socio-political jurisdiction maps where some landforms (such as rivers) have played a role in demarcation of boundaries, but which otherwise have limited relationship with environmental patterns (Fig. 1a), were also chosen to act as controls. Our results show that in the absence of strong selective filters, there is limited congruence between gene flow patterns and floristic or bioregional classifications.

Results

Identification of genetic lineages and the ubiquity of isolation by distance

Genomic data were obtained for a total of 1567 species-site combinations across 50 species. For all species, individuals inferred to be closely related based upon kinship values and clustering in principle coordinate analysis (PCoA) and SplitsTree analysis were sampled from the same site. Additionally, patterns of relatedness between sites were geographically sensible for most species, in that proximity was a major predictor of genetic similarity, and where this was not so, patterns of relatedness fit the existing understanding of the biogeography of the study region²¹. The number of genetic lineages identified for each species based upon population genetic analyses ranged from one to seven across the target species, with nine species that lacked clear genetic structure treated as a single genetic lineage (Supplementary Data 1). Most of the identified intra-specific genetic lineages were geographically discrete, sometimes with intergrade zones, but for two species geographically overlapping but genetically distinct lineages were identified. Conversely, four species showed geographic

structuring between coastal and non-coastal lineages that had similar latitudinal ranges, but were consistent in the habitats (coastal heath vs other habitats) they occupied.

Within the study area, many of the areas where genetic neighbourhoods abutted corresponded to known biogeographic features that act (or have done so in the past) as barriers to dispersal for some species (Fig. 2)²¹. Mantel tests for isolation by distance between sampling sites were significant at a p -value of < 0.01 for 45 of the 50 species (Supplementary Information). Those that did not show significance fell into two groups: species with very limited differentiation between any sites ($n = 3$), and those with high rates of inbreeding leading to high differentiation between geographically close sites ($n = 2$). The variation in strength of the signal of isolation by distance²² (Mantel test r values: 0.380–0.948) also suggests there is variation in the strength and potentially type of processes leading to differentiation across the landscape. Much of this variation in the strength of the isolation by distance (IBD) relationship is attributable to mating system and dispersal ability of each species, as seen by the strong effect of the rate of inbreeding (Supplementary Information).

The concordance between mapping schemes and genetic variation: Discriminant Analysis of Principal Components (DAPC)

Twenty-two DAPC analyses were performed for each of the 50 species. One DAPC analysis used k-mean clustering based on the genetic data with optimal cluster number chosen algorithmically. The remaining twenty-one included three DAPC analyses per mapping scheme ($n = 7$) with cluster number constrained to the number of mapping classes the species occurred in across the seven mapping schemes investigated. The three clustering methods used per mapping scheme were: algorithmically optimised sample assignment to clusters, assignment of samples to clusters corresponding to the mapping classes they were collected in, and randomised assignment for samples to clusters.

Across fifty species, optimised k-mean clustering showed the highest average proportion of successful reassignment values as returned by DAPC analyses. However, the optimised k-mean clustering result was not significantly different to the reassignment proportions of the DAPC performed on clusters of number equal to the mapping classes a species occurred in but with algorithmically optimised sample assignment to clusters for all mapping schemes excluding PCTs (Supplementary Data 2). For all mapping schemes averaged across species, the success rate of reassigning samples to the mapping classes from which they were collected in the DAPC was lower than that of clusters of the same quantity but with sample assignment algorithmically optimised. However, for Interim Biogeographic Regionalisation for Australia (IBRA) subregions, Local Government Areas (LGAs) and Plant Community Types (PCTs) this difference was not significant (Supplementary Data 2 and Fig. 3). The optimised clustering of schemes with fewer classes (IBRA regions, Vegetation Formations, and Local Land Services (LLSs)) showed higher successful reassignment than those with higher numbers of classes (PCTs and LGAs), showing a negative effect of increased cluster number in the optimised analyses.

For cluster assignment based upon mapping classes, of the seven mapping schemes, vegetation formation and vegetation class showed significantly lower reassignment success than the remaining five (Supplementary Data 2). While the remaining five were not significantly different in their average successful reassignment of samples to clusters, IBRA regions showed the highest average successful reassignment value, marginally higher than IBRA subregions. PCTs and LGAs performed similarly, and while having a lower average reassignment value than IBRA regions, the latter had a greater range of successful reassignment values resulting in a longer tail of low reassignment values for individual species, likely reflecting the changes in size of individual LGAs across the study region. We also see randomised sample clustering showed consistently lower successful

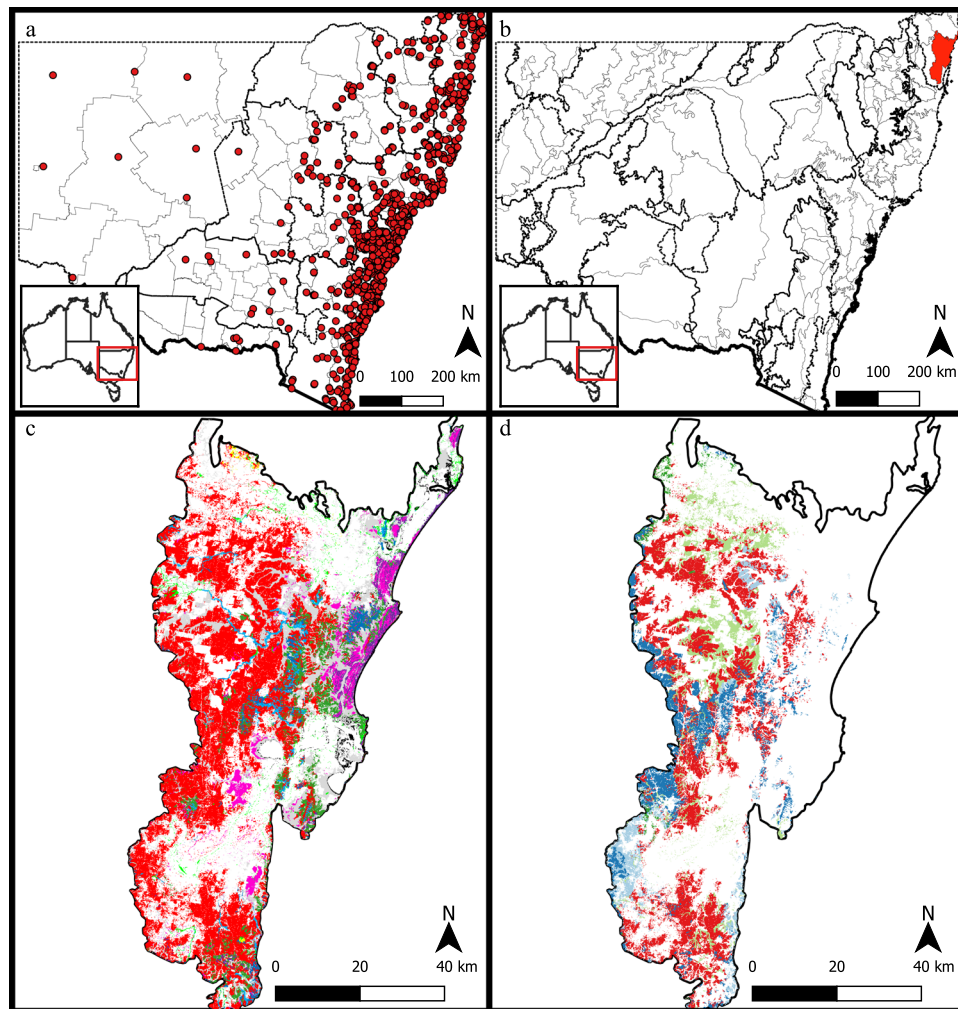


Fig. 1 | Maps of New South Wales, Australia showing sampling effort for genetic data generation and mapping schemes investigated in this study. Panel (a) shows the collecting localities of all 50 species (9606 samples in total – red points) analysed in this study over the eleven local land services within New South Wales⁴⁷ (Local Land Services, Trade and Investment NSW, Maitland NSW, Creative Commons Attribution 3.0 Australia <http://creativecommons.org/licenses/by/3.0/au/>, accessed from Research Data Australia [<https://researchdata.edu.au/local-land-services-nsw-20140205/1437504>], date accessed 2023-03-02) with thick dashed borders, and the local government areas⁴⁸ (State Government of NSW and Spatial Services (DCS) 2024, NSW Administrative Boundaries, Creative Commons Attribution 3.0 Australia <http://creativecommons.org/licenses/by/3.0/au/>, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/bb020282-59e4-4db5-98c9-9dc23cf0b4f5>], date accessed 2023-03-02) in thin dashed borders. Panel (b) shows the Interim Biogeographic Regionalisation for Australia (IBRA)⁹ (Commonwealth of Australia and Department of Climate Change, Energy, the Environment and Water 2024, Interim Biogeographic Regionalisation for Australia (IBRA), Version 7 (Regions), Creative

Commons Attribution 4.0 <https://creativecommons.org/licenses/by/4.0/>, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/interim-biogeographic-regionalisation-for-australia-ibra-version-7-regions>], date accessed 2016-04-14) regions (thick dashed borders) and subregions (thin dashed borders) with the Clarence Lowlands subregion (enlarged in panels c and d) shown in solid red. Also shown are examples of the mapping scale of (c) vegetation formations and (d) Plant Community Type (PCT)⁶⁹ (State Government of NSW and NSW Department of Climate Change, Energy, the Environment and Water 2022, NSW State Vegetation Type Map, Creative Commons Attribution 4.0 <https://creativecommons.org/licenses/by/4.0/>, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/nsw-state-vegetation-type-map>], date accessed 2023-06-09) in a single IBRA subregion, the Clarence Lowlands, where each colour represents a unique vegetation formation (11 formations shown by unique colours) or PCT (8 formations shown by unique colours) from a single vegetation formation (Dry Sclerophyll Forests (Shrub/grass sub-formation)).

reassignment values, suggesting all other clustering methods were better than random.

Correlation of genetic lineages and mapping schemes

When occurrence records of individual species located within 20 km of the collection location of genotyped samples were assigned to the corresponding genetic lineage, as expected, using survey records increased the number of classes individual genetic lineages were observed in across all mapping schemes compared to herbarium records (Fig. 4). Although this discrepancy was far larger for PCTs than other tested mapping schemes (vegetation formation and class were

not tested due to poor performance in DAPC analysis), in all cases it is likely due to the lower sampling effort needed for survey records leading to higher resolution data and a more comprehensive representation of species occupation of the landscape. Meanwhile, herbarium records were more limited in resolution, but higher in identification accuracy due to their nature as physical objects.

Table 1 shows the average number of mapping scheme classes that a single genetic lineage, as defined by records within 20 km of a genotyped sample, was recorded in across the 50 target species using either herbarium specimen records or survey records. Due to their greater geographic extent, single genetic lineages crossed far fewer

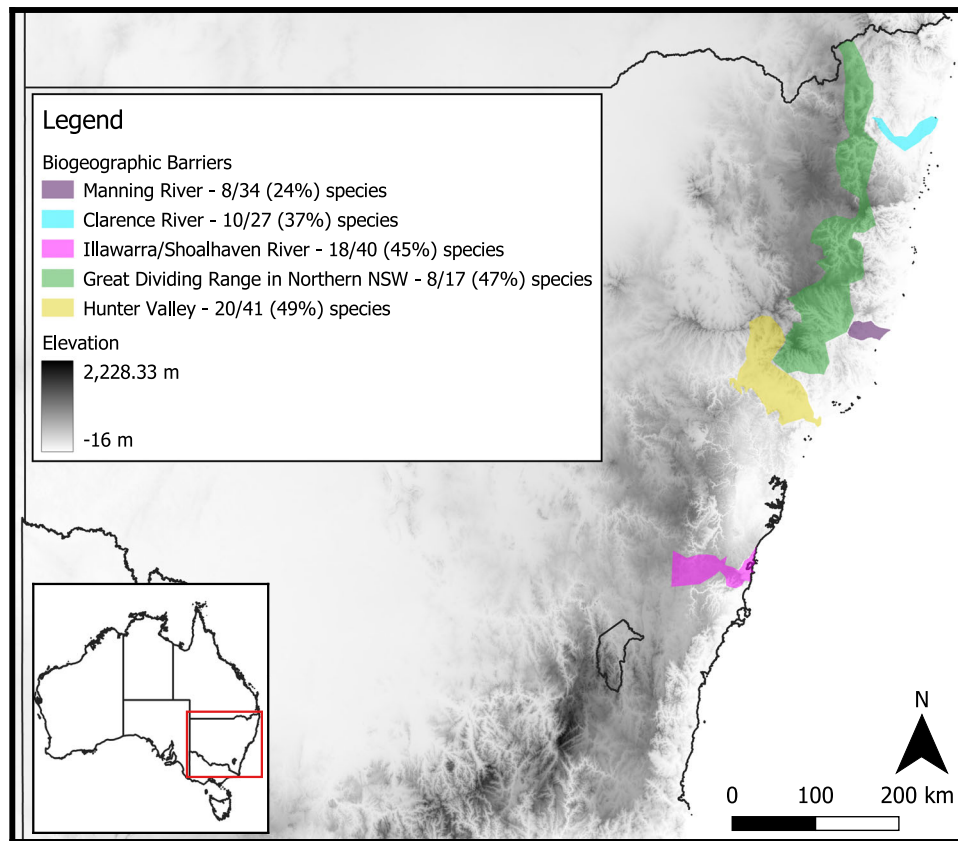


Fig. 2 | Map of major biogeographic barriers (coloured dotted regions) that corresponded to change in genetic neighbourhoods in multiple species in this study. The barriers are overlaid on an elevation map of the study region (grey scale: darker is higher elevations in metres) to highlight how the barriers align with

topographic features. The number of species with a change in genetic neighbourhood and the total number of studied species which occur across each barrier is indicated in the legend, as well as the percent this represents.

IBRA regions (~2.8) and LLSs (~3) than IBRA subregions (9 to 12 depending on record type). Conversely, single genetic lineages spanned many of the highly geographically constrained PCTs (25 to 80 depending on record type). Single genetic lineages spanned a moderate number of LGAs (13 to 18 depending on record type), however the scale of LGAs varies dramatically across the study region, making generalised interpretations at this scale difficult.

Interestingly, the lowest portion of mapping classes hosting multiple genetic lineages was observed for LGAs, with PCTs having a slightly higher value (Table 1). As can be expected, the large IBRA regions and LLSs showed the highest proportion of classes with multiple genetic lineages present (>20%), while the proportions for IBRA subregions was more like that of LGAs and PCTs (Table 1).

Discussion

This study, unique in its geographic reach and standardised multi-species approach, investigated the relationship between landscape-wide patterns of genetic variation and vegetation assembly patterns as described by vegetation and floristic classifications and bioregionalisations. We found limited correlation between the structuring of genetic diversity in 50 species distributed across large and heterogeneous geographic areas, and various descriptors of vegetation assemblage, indicating that floristic assembly patterns and classifications are not suitable predictors of intra-specific gene flow patterns and lineage divergence. Indeed, we found vegetation type classifications and bioregionalisations are no more predictive of intra-specific genetic patterns than socio-political boundaries in the LLS and LGA maps. This contrasts with findings of a previous European²³ study

which showed alignment between allele and species ranges. The difference between these two systems is that in the European context, both allelic and species ranges are responding to the same extreme environmental drivers, glaciation during ice ages, whereas in the regional context of this study, there is not a history of such extreme environmental changes²⁴. This allows for differences in the underlying processes of floristic and vegetation assemblage formation, and intra-specific patterns of genetic variation to be teased apart.

Our results show that although both processes can sometimes respond to similar environmental drivers²³, intra-species gene flow and dispersal patterns are not primarily responding to floristic and vegetation assembly patterns, something that had not been previously validated in real world systems². Previous studies have focused on the relationship between species diversity and intra-specific genetic diversity³, which is not the focus here^{2,23}. We show that regional sites with contrasting assemblages support more genetically similar populations than spatially disjunct but analogous vegetation types. Replicated sampling across 50 study species found that gene flow can be uninterrupted over large distances with IBD²² (observed for 90% of studied species) rather than vegetation and floristic assembly processes being a major driver of distribution-wide genetic patterns. Detectable genetic divergences were largely correlated to distributional breaks associated with biophysical gaps and recognised biogeographic barriers (Fig. 2)²¹. This explains why, within the context of the study system, the IBRA regions and subregions (contiguous, based upon both biotic and abiotic environmental factors), and the socio-political LLS and LGA maps (not directly based upon environmental factors but contiguous), are more closely tied to the species-specific

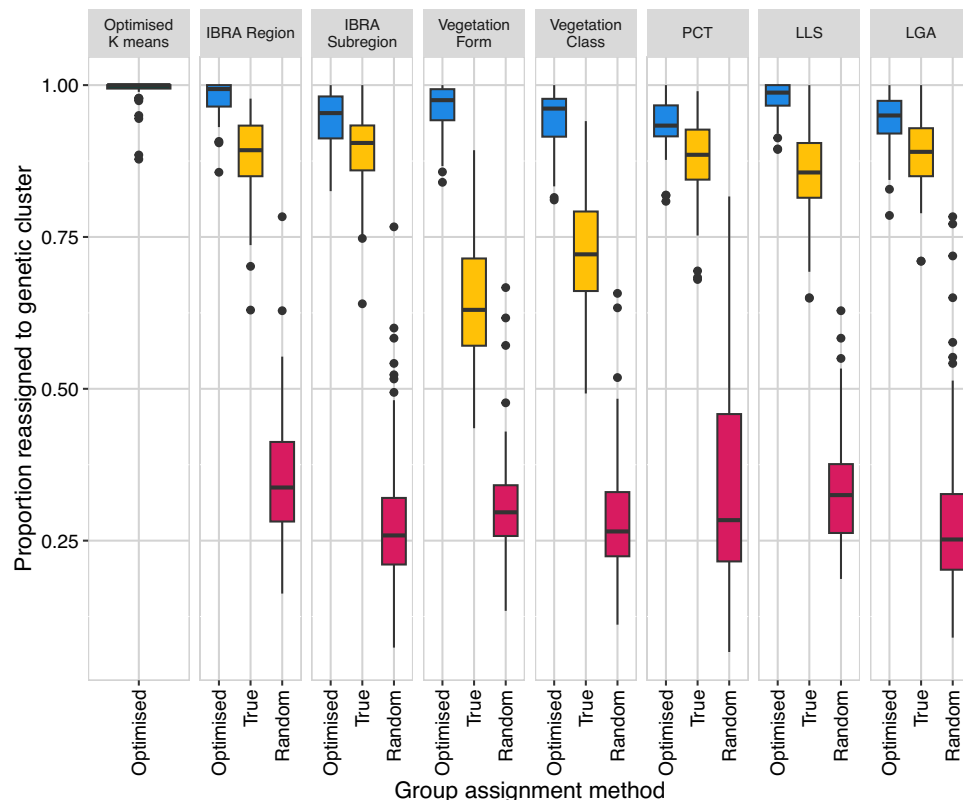


Fig. 3 | Average proportion of samples reassigned to various genetic clusters in discriminant analysis of principal components (DAPC) across 50 species. Box plots show median values with centre line, inter-quartile ranges with boxes, 1.5 interquartile range with whiskers, and outlying points as inline points. An unconstrained and optimised k means clustering analysis is shown on the left, while for seven mapping schemes the following analyses are presented: firstly, an approach where the number of clusters was constrained to match the number of mapping classes from which the species' samples originate, but with assignment to clusters unconstrained and optimised (blue); secondly, an approach where both the number of clusters and the sample assignment to clusters are constrained to align with

mapping classes from which samples were collected (yellow); and lastly, an approach where the cluster numbers are constrained to correspond to the number of mapping classes the samples are derived from, but with randomised assignment of samples to clusters (red). Optimised analyses performed better than mapping classes in all cases, although the latter still perform better than random. Mapping schemes tested include Interim Biogeographic Regionalisation for Australia (IBRA) regions and subregions, vegetation formations and classes, plant community types (PCT), Local Land Services (LLS) and Local Government Areas (LGA). Source data are provided as a Source Data file.

evolutionary data than the vegetation classification scheme. In this study system, boundaries of the former more often align with biogeographic features that limit species dispersal than those of the latter.

As our study focussed on species with large populations and geographic distributions, our findings may not be as pertinent to taxa with narrow or highly disjunct distributions. This is because strong barriers to gene flow between ecotypes could rapidly lead to the establishment of separate evolutionary trajectories and eventually speciation^{25–27}, as supported by the correlation of biogeographic barriers and genetic neighbourhood boundaries in this study (Fig. 2). Organisms other than plants, and with different distribution types (such as marine organisms²⁸), might also exhibit different associative patterns²⁷. However, the real-world implications of the scale of gene flow occurring across large distances comprising disparate vegetational assemblages, are not often applied in ecological restoration, where locally sourced material from within similar vegetation types or within management areas is often preferentially sought^{29,30}. Such practices are unlikely to mirror natural patterns of gene flow and are likely to exacerbate the deleterious consequences of increasing habitat fragmentation³¹. In addition, the increased pace of environmental change due to human activities increases the need for genetically diverse populations and large-scale movement of adaptable alleles across the landscape to ensure species survival in the long term³².

Our empirical evidence suggests that provenancing should not be limited to the local vegetation community, as this could unnecessarily

constrain the sustainability and evolutionary resilience of restoration plantings. Increasingly, targeted studies are highlighting how mixing source provenances in restoration targeted activities can result in greater adaptive potential and long-term sustainability of re-established populations^{33–35}. Although some of the theoretical discussions on provenance sourcing still revolve around the local vs non-local material debate, what is meant by 'local' might be the more important question^{18,36}. Even if the coarsest mapping scheme, IBRA regions, was to be used as a surrogate for local provenancing in restoration, single genetic lineages would often stretch across multiple mapping classes and therefore comprehensive collections within a single IBRA region might miss a lot of compatible diversity. Conversely, there remains a significant chance of mixing material from two or more distinct genetic lineages within these comprehensive collections. Overall, this suggests staying within a single mapping class for any mapping scheme does not guarantee that provenances are 'local', reproductively compatible (due to processes such as outbreeding depressions and phenology differences that may exist between genetic lineages) and capture a sufficiently high proportion of compatible genetic diversity.

Our finding that geographic distance and barriers are the primary drivers of genetic variation patterns in plant taxa (Fig. 2) corroborates previous findings in different regions of the world. For example, smaller multispecies studies in North America also identified geographic distance as the main genetic similarity predictor useful to select provenances to be potentially mixed during restoration^{37,38}.

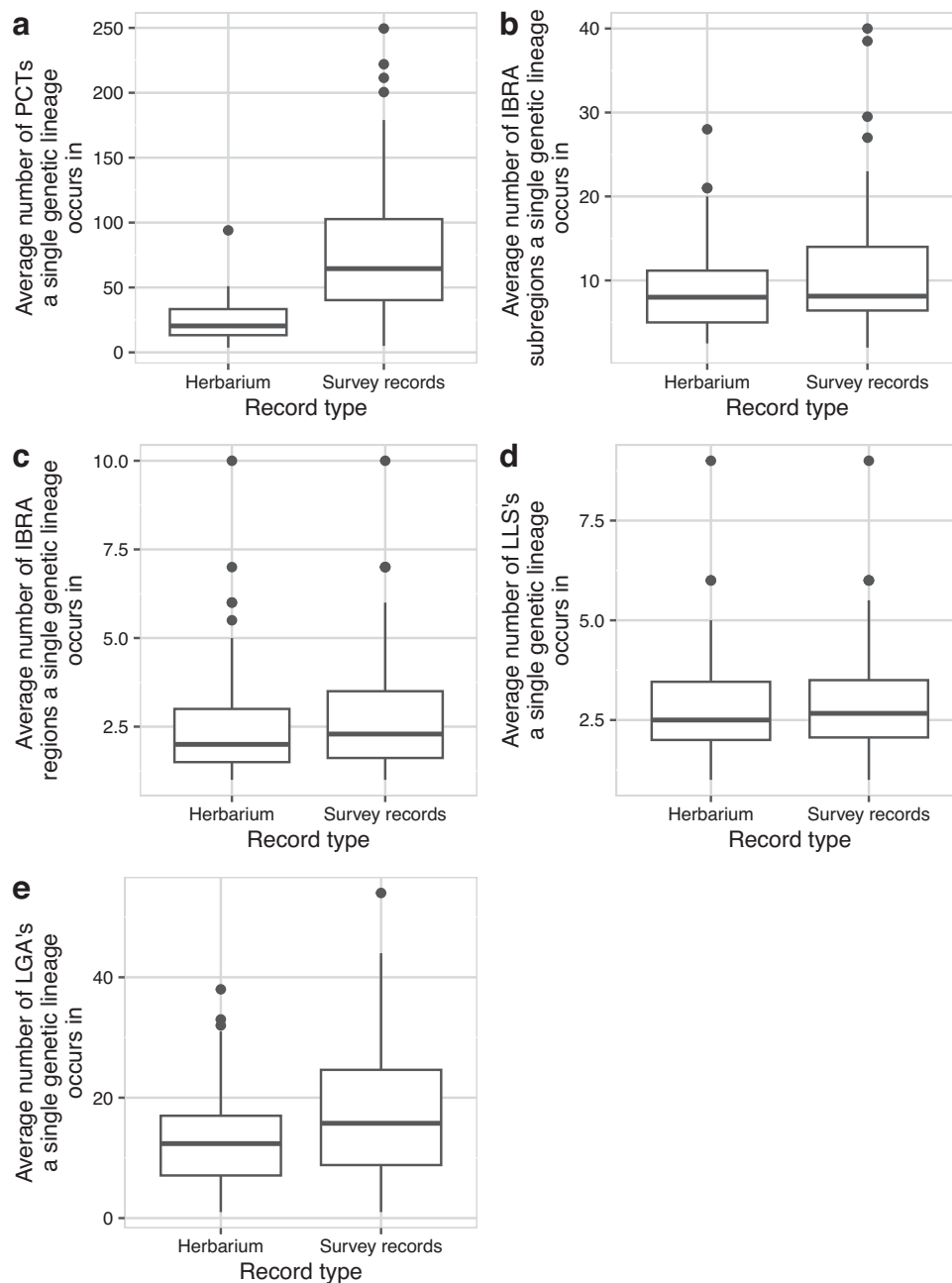


Fig. 4 | Boxplots showing the average number of landscape classification categories a single genetic lineage occurs in, across the distribution of 50 species, as derived from both herbarium specimen and survey records. Box plots show median values with centre line, inter-quartile ranges with boxes, 1.5 interquartile range with whiskers, and outlying points as inline points. The five mapping schemes investigated are (a) Plant Community Types (PCTs), (b) Interim

Biogeographic Regionalisation for Australia (IBRA) subregions, (c) IBRA Regions, (d) Local Land Services regions (LLSs), (e) Local Government Areas (LGAs). In all cases single genetic lineages occur across multiple mapping classes on average, with IBRA regions showing the lowest mapping classes for a single genetic lineage. Source data are provided as a Source Data file.

While we are not exploring the distribution of putatively adaptive genes, extensive adaptive research on Switchgrass^{39,40} concluded that adaptive ecotypes and genetic lineages should both be recognised as arising from between-population evolutionary distinctions caused by biogeographic barriers and isolation by distance effects. Further studies on fish⁴¹, birds⁴² and insects⁴³ have also found that neutral variation supports similar population genetic structuring as adaptive variation, although the structuring may be stronger for adaptive variation. Contrasting findings arose from a study of Swiss Stone Pine that showed contrasting patterns of genetic diversity between datasets containing neutral and putatively adaptive SNPs⁴⁴. This means we

cannot rule out that adaptive variation may produce different estimates of gene flow to the neutral SNPs used here, however these adaptive patterns are likely to vary significantly between species, further increasing the importance of species-specific genetic datasets rather than the use of proxies for gene flow.

Overall, this study provides evidence that within species gene flow and genetic structuring are independent and distinct from floristic assembly processes, and therefore one cannot be treated as a proxy for the other in the absence of species-specific data. Another outcome from this multispecies investigation is that taxa that occur as part of the same vegetation communities can have very different population

Table 1 | Key statistics from genetic lineage vs mapping scheme class analysis: the mean number and standard deviation of mapping classes a genetic lineage occurs in across 50 species, and the averaged percent of these classes that host multiple genetic lineages

Mapping scheme	Results when herbarium specimen records were used as the measure of species distribution		Results when survey records were used as the measure of species distribution	
	Mean number of mapping scheme classes in which a single genetic lineage occurs	Average percent of mapping scheme classes that host multiple genetic lineages	Mean number of mapping scheme classes in which a single genetic lineage occurs	Average percent of mapping scheme classes that host multiple genetic lineages
PCT	24.15 (sd = 15.56)	5.02% (sd = 5.29%)	79.64 (sd = 57.21)	10.08% (sd = 8.43%)
IBRA subregions	9.09 (sd = 5.65)	8.19% (sd = 9.86%)	11.55 (sd = 8.33)	8.57% (sd = 10.86%)
IBRA regions	2.69 (sd = 1.80)	24.35% (sd = 22.58%)	2.95 (sd = 1.88)	29.77% (sd = 22.96%)
LLS	2.95 (sd = 1.50)	31.63% (sd = 23.96%)	3.12 (sd = 1.53)	32.28% (sd = 25.08%)
LGA	13.68 (sd = 8.79)	5.23% (sd = 7.89%)	18.11 (sd = 11.82)	6.29% (sd = 8.01%)

Data is presented for both herbarium specimen and survey records collected within 20 km of a sampling site. Results for five mapping schemes are shown: plant community types (PCTs), Interim Biogeographic Regionalisation for Australia (IBRA) regions and subregions, Local Land Services (LLSs) and local government areas (LGAs). Source data are provided as a Source Data file.

genetics patterns, matching the findings of a previous study from Europe⁴⁷, which is directly relevant to restoration and management practices. A key future line of enquiry utilising large, multispecies genomic datasets as presented here, is to determine the role of isolation by environment in predicting intra-specific gene flow patterns and looking for possible environmental drivers steering these intra-specific processes. Exploring the relationship between patterns of adaptive and neutral genetic structuring may further clarify the application of genetic markers in in natural areas management. By focusing restoration genetic studies on highly abundant and ecological significant species that are common targets in environmental restoration practices⁴⁵, it is possible to largely eliminate the need for interpretative proxies and generalisations⁴⁶.

Methods

Geographic context, target species and sampling strategy

Fifty plant species were sampled across their distribution in NSW (Fig. 1a), a large (801,150 km² land area), environmentally heterogeneous area that includes the majority of biomes present on the Australian continent, with the aim of fully sampling the distribution of each species within the state. The species targeted for this study (listed in Supplementary Data 1) form part of the Restore and Renew restoration genetics program funded by the NSW government¹⁸ and represent life forms ranging from grasses and herbs to woody shrubs and trees across eighteen plant families representing different distributional ranges and environmental tolerances. The number of sites sampled per species reflects these distributional differences, ranging from 7 (35 individuals) to 74 (370 individuals – Supplementary Data 1).

In short, leaf material from five to twelve individuals of each species were collected at each sampling site, with an aim to have 10–100 m between sampled individuals at a site to limit the likelihood of sampling siblings or clonal individuals. One herbarium voucher per site, lodged at the National Herbarium of NSW, was collected from a sampled individual for identification confirmation. Data for the mapping schemes of interest were also downloaded, with the Interim Biogeographic Regionalisation for Australia (IBRA)⁹ (Commonwealth of Australia and Department of Climate Change, Energy, the Environment and Water 2024, Interim Biogeographic Regionalisation for Australia (IBRA), Version 7 (Regions), Creative Commons Attribution 4.0 <https://creativecommons.org/licenses/by/4.0/>), accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/interim-biogeographic-regionalisation-for-australia-ibra-version-7-regions>], date accessed 2016-04-14), Local Land Service boundaries⁴⁷ (Local Land Services, Trade and Investment NSW, Maitland NSW, Creative Commons Attribution 3.0 Australia <http://creativecommons.org/licenses/by/3.0/au/>), accessed from Research Data Australia [<https://researchdata.edu.au/local-land-services-nsw-20140205/1437504>], date accessed 2023-03-02) and Local Government Areas⁴⁸ (State Government of NSW and Spatial Services (DCS) 2024, NSW Administrative Boundaries, Creative Commons Attribution 3.0 Australia <http://creativecommons.org/licenses/by/3.0/au/>), accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/bb020282-59e4-4db5-98c9-9dc23cf0b4f5>], date accessed 2023-03-02) downloaded as shapefiles, and the NSW vegetation classification framework⁴⁹ (State Government of NSW and NSW Department of Climate Change, Energy, the Environment and Water 2022, NSW State Vegetation Type Map, Creative Commons Attribution 4.0 <https://creativecommons.org/licenses/by/4.0/>), accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/nsw-state-vegetation-type-map>], date accessed 2023-06-09) as a raster layer.

Data generation, filtering, and identification of genetic lineages

Data generation and analyses followed the Restore and Renew pipeline¹⁸. Small pieces (~ 0.5 cm × 2 cm) of leaf material from each

sample were placed into microtubes and sent to Diversity Array Technologies Pty Ltd (Canberra) for DArTseq data generation^{50–52}, an enzymatic genomic complexity reduction platform with the resultant data provided in the form of single nucleotide polymorphisms (SNPs) that are either anonymous, as in the case of this study, or mapped to a reference genome. Using the *dartR*⁵³ package and the *Restore* and *Renew* analysis pipeline^{48,54} in *R*⁵⁵, the DArT SNP files were used to create datasets for each species containing all samples with data for at least 50% of loci and subsampling to include only five samples per site, excluding sites with less than five successfully sequenced samples from the dataset. However, for four species with high rates of selfing and vegetative spread, sites with less than 5 distinct genets were included in clustering analyses as if these were excluded, the few remaining sites were not geographically representative. Datasets were further filtered to include only one SNP per sequenced locus (~ 75 bp restriction site associated genome region) with representation of > 80% of samples and a reproducibility score of > 0.96. Samples that were misidentified, of hybrid origin or suspected of being planted, were removed based on clustering in PCoA and *SplitsTree* analyses and pairwise F_{st} values in a preliminary screening. For all species, population genetic analyses undertaken included:

- Calculation of pairwise kinship values between samples within sites using the PLINK method as implemented in the *snpgdsIBD-MoM* function from the package *SNPRelate*⁵⁶ to remove clonal samples using a threshold kinship value of 0.4.
- The calculation of population genetic statistics including expected heterozygosity (H_e), observed heterozygosity (H_o), allelic richness (Ar), number of private alleles (Ap), and inbreeding coefficient (F_{is}) values on a site-by-site basis using the *diversity* package⁵⁷ to build an understanding of species level of genetic diversity and inbreeding.
- A principal coordinate analysis (PCoA) using the *pcoa* function from the *ape* package⁵⁸ to investigate how variation and clustering in the dataset related to sampling sites.
- Phylogenetic network analysis in *SplitsTree* version 4.17.2⁵⁹ to investigate sample clustering and the relationships between sampled sites.
- A sNMF analysis⁶⁰ to investigate the number and identity of ancestral lineages of each species.
- Calculation of inter-site F_{st} values using the *SNPRelate* package⁵⁶ to investigate level of population divergence and use in a Mantel test for isolation by distance between all pairs of sampling sites utilising the *Vegan* package⁶¹.

Major genetic lineages were identified by taking the conservative approach of choosing the lowest number of lineages where there was clear genetic differentiation into groupings based on all analyses including *SplitsTree* network, PCoA, sNMF and F_{st} results, whether these groupings were geographically discrete or not. In most cases genetic lineages used in further analyses were congruent with sNMF ancestral groupings with the lowest cross-entropy value, or in the case that there were several values for ancestral populations with similarly low cross-entropy values, the lowest number of ancestral populations in the group was chosen. In some cases, high levels of inbreeding lead to an inflation in the optimal number of ancestral lineages, in which case genetic lineages were identified based primarily on clustering in PCoA and network analyses which were less affected by this. Supplementary Data 1 shows the number of samples, sampling sites and SNPs in the final dataset for all species.

DAPC analysis

The *find.cluster* function from *adegenet*^{62,63} was used to determine optimal number of clusters to be used in discriminant analysis of principal components (DAPC)⁶⁴. After testing increasing numbers of iterations and random starts to test the stability of BIC values, ten

million iterations and one thousand random starts were chosen for use in analyses on all species. Optimal numbers of clusters were chosen automatically using the *diffNgroup* method on the BIC statistic. DAPC analysis was then performed using this clustering scheme, as well as IBRA regions, IBRA subregions, vegetation form, vegetation class, PCT, LLS and LGA groupings. To test whether the variation in resulting clustering efficiency was purely due to cluster numbers, we used the *find.cluster* function to find optimal cluster membership when the number of clusters was constrained as equal to the number of mapping classes a species was collected from for each of the seven mapping schemes and ran a DAPC analyses on each of these clustering schemes. As a control, sites were randomly assigned to a number of clusters equal to other analyses, followed by a DAPC analyses on the random clustering. In all DAPC analyses, the number of principal components (PCs) retained was determined by initially running a DAPC where the number of PCs equalled the number of clusters minus one. The *optm.a.score* function was then used to choose an optimal number of PCs to retain based upon this initial DAPC, limiting the maximum possible number of PCs to the number of clusters being tested, in accordance with the suggestions of Thia (2022)⁶⁵. The *assign.prop* value provided in the output from the DAPCs, which represents the averaged proportion of individuals successfully reassigned to their a priori cluster⁶⁴, was recorded for each DAPC and used as a measure of the fit of the clustering to the data to compare clustering schemes. Dunn's test of multiple comparisons as implemented by the *dunn.test* R package⁶⁶ was then used to investigate differences in the average successful reassignment proportion between mapping and clustering schemes, employing a Holm correction for multiple comparisons (Supplementary Data 2).

Representation of environmental classification across genetic lineages

To investigate the relationship between different mapping schemes and the distribution of genetic lineages, occurrence records of target species were downloaded from the Atlas of Living Australia (ALA) using an in-house R package that expands upon the *galah* package⁶⁷. Records were filtered to remove duplicates, cultivated specimens and those with visibly incorrect location information. Two datasets were created:

1. Records with associated herbarium specimens.
2. Survey records where no physical specimen was collected.

Dataset 1 had higher accuracy in the taxonomic identification of each record, however the accuracy of location data is generally lower, particularly for records pre-dating GPS. On the other hand, the location data for the second data set is more accurate while identification cannot be verified and may be less reliable. These surveys also represent more remote areas less frequently visited.

For each species, a polygon shapefile containing 20 km circular buffers around each of the sites where genetic samples were collected from was created in QGIS version 3.22.1⁶⁸ and each polygon was assigned to a single genetic lineage corresponding to the genetic sample it was centred on. The distance of 20 km was chosen as a standard under the assumption that, based upon the results of our initial species population genetic analyses, gene flow over this distance is significant and therefore populations within this radius have a high likelihood of belonging to the same genetic lineage. These shapefiles were then read into R along with data for the four mapping schemes. Then for each species, ALA records that fell within one or more of the 20 km buffers were assigned to the relevant genetic lineage/s and for each of these records the IBRA region, IBRA subregion, PCT, LLS and LGA it occurred in was recorded. Vegetation formation and vegetation class from the NSW vegetation classification framework were not included in this analysis due to their poor fit to the data in the DAPC analyses and less contiguous nature. For each species, genetic lineages

were summarised to arrive at the number of IBRA regions, IBRA sub-regions, PCTs, LLSs and LGAs that lineage occurred in. The average number for each genetic lineage per species was then calculated and plotted in Fig. 4. The proportion of classes in each scheme which occurred in two or more genetic lineages for any species was also calculated.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sample collection data and SNP data generated in this study have been deposited in the University of Queensland's eSpace data repository under accession code 10.48610/5c76de2 (<https://doi.org/10.48610/5c76de2>). Raw sequencing data is held by Diversity Array Technologies Pty Ltd, and use of this data can be organised by contacting the authors. Source data for Figs. 3 and 4 and Table 1 are provided as a Source Data file. The individual species population genetic information generated in this study are provided in the Supplementary Information file. Source data are provided with this paper.

Code availability

All analytical code used in this study is part of freely available programs as cited in the text. Where in-house R pipelines are mentioned, these refer only to scripts built on servers that implement the cited analytical code as part of the Restore and Renew program.

References

- Chesson, P. Updates on mechanisms of maintenance of species diversity. *J. Ecol.* **106**, 1773–1794 (2018).
- Vellend, M. et al. Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Mol. Ecol.* **23**, 2890–2901 (2014).
- Vellend, M. & Geber, M. A. Connections between species diversity and genetic diversity. *Ecol. Lett.* **8**, 767–781 (2005).
- Taberlet, P. et al. Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecol. Lett.* **15**, 1439–1448 (2012).
- Zobel, M. The species pool concept as a framework for studying patterns of plant diversity. *J. Veg. Sci.* **27**, 8–18 (2016).
- Keith D. *Ocean shores to desert dunes: the native vegetation of New South Wales and the ACT*. Department of Environment and Conservation (NSW) (2004).
- Davies C. E., Moss D. & Hill M. O. EUNIS habitat classification revised 2004. In: *Report to European Environment agency-European topic centre on Nature Protection and biodiversity* (2004).
- Rehfeldt, G. E., Crookston, N. L., Sáenz-Romero, C. & Campbell, E. M. North American vegetation model for land-use planning in a changing climate: a solution to large classification problems. *Ecol. Appl.* **22**, 119–141 (2012).
- Australian Government Department of Climate Change, Energy, the Environment and Water. Interim Biogeographic Regionalisation for Australia (IBRA) Version 7 (Subregions), accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/8e242336-7d10-4630-ae81-e1b6e7464f3c>] (2023).
- Franklin, J. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* **19**, 474–499 (1995).
- Wagner, H. H. & Fortin, M.-J. Spatial analysis of landscapes: concepts and statistics. *Ecology* **86**, 1975–1987 (2005).
- McIntosh, R. P. Plant communities: recent research suggests that they form units in a vegetation continuum rather than discrete classes. *Science* **128**, 115–120 (1958).
- De Cáceres, M. & Wiser, S. K. Towards consistency in vegetation classification. *J. Vegetation Sci.* **23**, 387–393 (2012).
- Harris, J. A., Hobbs, R. J., Higgs, E. & Aronson, J. Ecological restoration and global climate change. *Restor. Ecol.* **14**, 170–176 (2006).
- Suding, K. et al. Committing to ecological restoration. *Science* **348**, 638–640 (2015).
- Thomas, E. et al. Genetic considerations in ecosystem restoration using native tree species. *Ecol. Manag.* **333**, 66–75 (2014).
- Durka, W. et al. Genetic differentiation within multiple common grassland plants supports seed transfer zones for ecological restoration. *J. Appl. Ecol.* **54**, 116–126 (2017).
- Rossetto, M. et al. Restore and Renew: a genomics-era framework for species provenance delimitation. *Restor. Ecol.* **27**, 538–548 (2019).
- Department of Planning and Environment, Government of New South Wales. A revised classification of plant communities of eastern New South Wales. Department of Planning and Environment, Government of New South Wales (2022).
- Thackway, R. & Cresswell, I. D. *An interim biogeographic regionalisation for Australia: a framework for establishing the national system of reserves*, Version 4.0. edn. Australian Nature Conservation Agency (1995).
- Bryant, L. & Krosch, M. Lines in the land: a review of evidence for eastern Australia's major biogeographical barriers to closed forest taxa. *Biol. J. Linn. Soc.* **119**, 238–264 (2016).
- Wright, S. Isolation by distance. *Genetics* **28**, 114–138 (1943).
- Thiel-Egenter, C. et al. Break zones in the distributions of alleles and species in alpine plants. *J. Biogeogr.* **38**, 772–782 (2011).
- Byrne, M. Evidence for multiple refugia at different time scales during Pleistocene climatic oscillations in southern Australia inferred from phylogeography. *Quat. Sci. Rev.* **27**, 2576–2585 (2008).
- Richards, T. J. & Ortiz-Barrientos, D. Immigrant inviability produces a strong barrier to gene flow between parapatric ecotypes of *Senecio latus*. *Evolution* **70**, 1239–1248 (2016).
- Sianta, S. A. & Kay, K. M. Parallel evolution of phenological isolation across the speciation continuum in serpentine-adapted annual wildflowers. *Proc. R. Soc. B: Biol. Sci.* **288**, 20203076 (2021).
- Stronen, A. V., Norman, A. J., Vander Wal, E. & Paquet, P. C. The relevance of genetic structure in ecotype designation and conservation management. *Evolut. Appl.* **15**, 185–202 (2022).
- Pena, R. R. & Colgan, D. J. Does marine bioregionalisation provide a framework for the conservation of genetic structure? *Regional Stud. Mar. Sci.* **40**, 101505 (2020).
- Baer, S. G. et al. No effect of seed source on multiple aspects of ecosystem functioning during ecological restoration: cultivars compared to local ecotypes of dominant grasses. *Evolut. Appl.* **7**, 323–335 (2014).
- McMullen, C. M. Limits to local sourcing in herbaceous plant restoration. *Ecol. Restor.* **40**, 64–69 (2022).
- Frankham, R. et al. *Genetic management of fragmented animal and plant populations*, First edition. edn. Oxford University Press (2017).
- Avik, T. & Helm, A. Restoration of plant species and genetic diversity depends on landscape-scale dispersal. *Restor. Ecol.* **26**, S92–S102 (2018).
- St. Clair, A. B., Dunwiddie, P. W., Fant, J. B., Kaye, T. N. & Kramer, A. T. Mixing source populations increases genetic diversity of restored rare plant populations. *Restor. Ecol.* **28**, 583–593 (2020).
- Zeng, X. & Fischer, G. A. Using multiple seedlots in restoration planting enhances genetic diversity compared to natural regeneration in fragmented tropical forests. *Ecol. Manag.* **482**, 118819 (2021).
- Höfner, J. et al. Populations restored using regional seed are genetically diverse and similar to natural populations in the region. *J. Appl. Ecol.* **59**, 2234–2244 (2022).

36. McKay, J. K., Christian, C. E., Harrison, S. & Rice, K. J. “How local is local?”—A review of practical and conceptual issues in the genetics of restoration. *Restor. Ecol.* **13**, 432–440 (2005).
37. Massatti, R., Shriver, R. K., Winkler, D. E., Richardson, B. A. & Bradford, J. B. Assessment of population genetics and climatic variability can refine climate-informed seed transfer guidelines. *Restor. Ecol.* **28**, 485–493 (2020).
38. Shryock, D. F. et al. Landscape genetic approaches to guide native plant restoration in the Mojave Desert. *Ecol. Appl.* **27**, 429–445 (2017).
39. Lovell, J. T. et al. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**, 438–444 (2021).
40. Zhang, Y. et al. Natural hybrids and gene flow between upland and lowland switchgrass. *Crop Sci.* **51**, 2626–2641 (2011).
41. Candy, J. R. et al. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Mol. Ecol. Resour.* **15**, 1421–1434 (2015).
42. Peters, J. L. et al. Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). *Biol. Conserv.* **203**, 272–281 (2016).
43. Batista, P. D., Janes, J. K., Boone, C. K., Murray, B. W. & Sperling, F. A. H. Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). *Ecol. Evol.* **6**, 6292–6300 (2016).
44. Dauphin, B. et al. Disentangling the effects of geographic peripherality and habitat suitability on neutral and adaptive genetic variation in Swiss stone pine. *Mol. Ecol.* **29**, 1972–1989 (2020).
45. Cooper, D. L. M. et al. Consistent patterns of common species across tropical tree communities. *Nature* **625**, 728–734 (2024).
46. Hoban, S. et al. Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. *Biol. Rev.* **97**, 1511–1538 (2022).
47. Local Land Services, Trade and Investment NSW. Local Land Services Spatial Layer NSW 20140205. In: *Bioregional Assessment Source Dataset* (2014).
48. State Government of New South Wales and Spatial Services (DCS). NSW Administrative Boundaries, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/bb020282-59e4-4db5-98c9-9dc23cf0b4f5>] (2024).
49. Department of Planning and Environment, Government of New South Wales. BioNet Plant Community Type data. (ed Department of Planning and Environment, Government of New South Wales) (2022).
50. Cruz, V. M. V., Kilian, A. & Dierig, D. A. Development of DaRT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop *Lesquerella* and related species. *PLOS ONE* **8**, e64062 (2013).
51. Kilian, A. et al. Diversity Arrays Technology: a generic genome profiling technology on open platforms. *Data Prod. Anal. Popul. Genomics* **888**, 67–89 (2012).
52. Sansaloni, C. et al. Diversity Arrays Technology (DaRT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc.* **5**, P54 (2011).
53. Gruber, B., Unmack, P. J., Berry, O. F. & Georges, A. DART An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* **18**, 691–699 (2018).
54. Bragg J. G. RRtools: Filtering, conversion, analysis of SNP genotype data. R package version 0.1 edn (2023).
55. R. Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2016).
56. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
57. Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. & Prodöhl, P. A. *diveRsity*: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788 (2013).
58. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
59. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
60. Frichot, E. & Francois, O. LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
61. Oksanen J. et al. *vegan*: Community Ecology Package. R package version 2.5-6 (2019).
62. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
63. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
64. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94–94 (2010).
65. Thia, J. A. Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Mol. Ecol. Resour.* **23**, 523–538 (2023).
66. Dinno A. dunn.test: Dunn’s test of multiple comparisons using rank sums. R package version 1.3.5 edn (2022).
67. Westgate, M., Stevenson, M., Kellie, D. & Newman, P. galah: Biodiversity data from the GBIF node network (2023).
68. QGIS.org. QGIS Geographic Information System. QGIS Association (2023).
69. State Government of New South Wales and New South Wales Department of Climate Change, Energy, the Environment and Water. NSW State Vegetation Type Map, accessed from The Sharing and Enabling Environmental Data Portal [<https://datasets.seed.nsw.gov.au/dataset/95437fbd-2ef7-44df-8579-d7a64402d42d>] (2022).

Acknowledgements

We acknowledge the Traditional Custodians of the land on which the plant species in this study are found on and pay respects to Elders past and present. We acknowledge the many collectors who facilitated the sampling of plant material used in this study, in particular Daniel Clarke (Arcane Botanica Pty Ltd), Joel Cohen (Tondoon Botanic Gardens), Robert Kooyman and Tricia Hogbin (Botanic Gardens of Sydney). We acknowledge those who have generated and curated the data used in this study including Patricia Lu-Irving (Royal Botanic Gardens Sydney), Susan Rutherford (Jiangsu University), Samantha Yap (Royal Botanic Gardens Sydney), Hannah MacPherson (National Herbarium of NSW), Kit King (Biosis Pty Ltd) and Karina Guo (Royal Botanic Gardens Sydney). This study was undertaken through non-competitive funding from the Royal Botanic Gardens and Domain Trust (P.S.F., R.D., M.R., M.M.M. and J.G.B.), New South Wales Department of Planning, Infrastructure, and the Environment (M.R.) and HSBC Bank Australia (M.R.), and a grant from the New South Wales Environmental Trust (Grant number: 2016/RD/0084; M.M.M.).

Author contributions

P.S.F., R.D., M.M.M., J.G.B., and M.R. contributed equally to conceiving the project, developing research questions, and interpreting results. M.R., M.M.M., and J.G.B. established the Restore and Renew program through which genetic data and population genetic analytical pipelines were generated. P.S.F. performed all analyses. P.S.F. and M.R. wrote the manuscript with input from R.D., M.M.M., and J.G.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54930-7>.

Correspondence and requests for materials should be addressed to Patrick S. Fahey or Maurizio Rossetto.

Peer review information *Nature Communications* thanks Felix Gugerli and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024