RESTORATION ECOLOGY
The Journal of the Society for Ecological Restoration
SER SOCIETY FOR ECOLOGICAL RESTORATION

RESEARCH ARTICLE

# Restore and Renew: a genomics-era framework for species provenance delimitation

Maurizio Rossetto[1,2], Jason Bragg[1], Andrzej Kilian[3], Hannah McPherson[1], Marlien van der Merwe[1], Peter D. Wilson[1]

**Here we present "Restore and Renew," a replicable framework for gathering and interpreting evolutionary, ecological, and genomic data in support of restoration practices. In an era of rapid climatic change and continuous widespread clearing, revegetation projects need to focus on producing resilient and long-term self-sustaining populations. Restore and Renew expands current knowledge of genetic provenance via genome-scan data, environmental niche modeling (ENM), and site-specific climate information. The sampling strategy is to obtain leaf tissue representing the distributions of over 100 species commonly used in restoration. We apply generalized dissimilarity modeling to genome-wide single nucleotide polymorphism datasets from hundreds of samples. Species-specific local provenances are obtained using a model that represents observed patterns of genetic variation across the landscape. Climate modeling is implemented to interpret genetic provenance boundaries in the context of current and future climatic conditions at the specified site. Results are presented in an easy-to-use webtool (www.restore-and-renew.org.au), where the user simply selects their site of interest and a target species to obtain the size and distribution of local genetic provenance. Although Restore and Renew is not prescriptive, it allows restoration practitioners to make informed decisions on where to source material from, to fulfill their restoration scenario of choice. Two examples, *Westringia fruticosa* and *Acacia suaveolens*, are presented to demonstrate how the analytical pipeline responds to different ecological and evolutionary patterns. The webtool has multiple applications for biodiversity management and will continue to evolve with new species and analytical/interpretative outputs.**

**Key words: ecological restoration, genetic provenance, Restore and Renew, single-nucleotide polymorphisms, webtool**

### Implications for Practice

- Restore and Renew establishes a novel framework for gathering, across multiple species, the climatic and genetic provenance information needed to execute restoration scenarios, initiate replicable experimental trials, and develop monitoring strategies. The framework is replicable globally across any vegetation type.
- Restore and Renew sets a precedent for large-scale, empirically based projects that support applied restoration activities by providing an easy to access and easy to interpret webtool (www.restore-and-renew.org.au) that can be integrated with other data-gathering schemes and updated with future developments.
- Access to large, dedicated empirical datasets that originate from technological and analytical developments will invigorate experimentally based ecological restoration research and diminish reliance on broad generalizations.

## Introduction

Environmental pressures such as clearing and habitat fragmentation impact plant communities by reducing the size and diversity of populations and preventing landscape-level connectivity (Frankham et al. 2017). To remediate the loss of vegetation cover, there has been a growing emphasis on habitat restoration through revegetation. These increasingly large-scale projects (e.g. www.bonnchallenge.org) need to be supported by suitable ecological and evolutionary information to be successful and mitigate anthropogenic impacts in the long term (Falk et al. 2006; Kettenring et al. 2014). So far, this information has been available only for a small number of species, and restoration practitioners have mostly relied on generalized scenarios that remain difficult to evaluate without reference data. Recent technological and analytical advances in genomic research are generating new opportunities for broader evolutionary explorations in restoration genetics (Williams et al. 2014).

Restoration and revegetation activities have been estimated at around US$2 trillion per annum globally (Prober et al. 2015).

Author contributions: MR devised the concept and wrote the manuscript; JB, PDW, AK, HM, MvM developed analytical pipelines and contributed to the writing.

[1]National Herbarium of New South Wales, Royal Botanic Garden Sydney, Mrs Macquaries Road, Sydney, NSW 2000, Australia
[2]Address correspondence to M. Rossetto, email maurizio.rossetto@rbgsyd.nsw.gov.au
[3]Diversity Arrays Technology, University of Canberra, Bruce, ACT 2617, Australia

*© 2018 Royal Botanic Garden and Domain Trust. Restoration Ecology published by Wiley Periodicals, Inc. on behalf of Society for Ecological Restoration*
*This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.*
doi: 10.1111/rec.12898
Supporting information at:
http://onlinelibrary.wiley.com/doi/10.1111/rec.12898/suppinfo

Considering the scale involved and the resources invested, it is critical that the goals of ecological restoration, such as reestablishing self-sustaining populations, ameliorating degraded ecosystems, and reinstating ecosystem services (Kettenring et al. 2014; McDonald et al. 2016), are achievable. For reestablished vegetation to be resilient and ecologically adapted (and adaptable), it needs to possess sufficient evolutionary potential to respond to selective filtering now and into the future (Lesica & Allendorf 1999). This is particularly true for anthropogenic climate change scenarios predicting altered timing of life history traits, shifts in geographical ranges, and modifications in competitive interactions (Sgro et al. 2011). These additional stressors and the following loss of genetic variation are likely to intensify the risk of localized extinctions (Pachauri et al. 2014). Restoration practices that maximize genetic diversity can therefore have a positive impact on ecosystem resilience and biodiversity management in general (Breed et al. 2013; Mijangos et al. 2015).

Within this context, the provenance delimitation discourse has shifted relatively recently from predominantly sampling from local sources ('local is best') to considering representation from a broader genetic basis (Broadhurst et al. 2008). This is because while excessive use of nonlocal material could potentially result in outbreeding depression (Fenster & Galloway 2000), increasing evidence highlights the importance and immediacy that inbreeding depression can have on populations with a narrow genetic spectrum (Frankham et al. 2011; Hedrick & Garcia-Dorado 2016; Szűcs et al. 2017). Consequently, restoration approaches that include material sourced more openly are increasingly considered as more suitable for achieving self-sustaining restoration targets.

As evolutionary biologists are still exploring the potential of populations and species to adapt through selection, plasticity, and/or gene flow, the relative risks of inbreeding or outbreeding depression for most species remain empirically untested (Weeks et al. 2011). In response to such uncertainties, a range of carefully crafted strategies for sourcing germplasm in restoration and revegetation projects has been proposed. These strategies range from "local sourcing" (McKay et al. 2005) to broader genomic representations including "composite" (Broadhurst et al. 2008), "predictive" (Crowe & Parker 2008), "admixture" (Breed et al. 2013), and "climate adjusted" (Prober et al. 2015) provenancing. Although these strategies (and the many others that have arisen in the last decade) are conceptually sound, they are difficult to execute in the absence of the empirical data necessary to identify those provenance boundaries on which they are based. Where does "local" end, and how do those boundaries change across the distribution of a species or between species? How are future climatic conditions likely to impact the current distribution of target species? If we cannot answer these questions in a replicable manner, we cannot develop experimentation to test the relative merits of the various strategies, monitor long-term success, nor consider the need for corrective measures. Consequently, decision-making about evolutionary implications is often reliant on inference and generalizations that can jeopardize the planned outcome (Rice & Emery 2003).

Despite the paucity of empirical data supporting provenance-sourcing decisions, restoration-granting agencies often recommend or even require specific sourcing strategies. "All-purpose" seed sourcing guidelines have been developed to support provenancing objectives (in Australia a commonly used example is the FloraBank Guideline 5; http://www.florabank.org.au). These have broad appeal because they set seed sourcing and provenancing criteria for a range of circumstance and plant-type combinations, but are constrained by evolutionary data availability. This is an important point because meta-analyses of plant genetic studies suggest that generalized patterns can only be validated across a limited array of functional and geographical attributes (Krauss & Koch 2004; Rossetto & Kooyman 2011; Broadhurst et al. 2017). Consequently, in the absence of directly relevant information, the definition of "local provenance" can range from sampling within a few hundred meters, to across a Bioregion (Hancock & Hughes 2012; Perring et al. 2015).

Although defining what constitutes "local" remains elusive without dedicated genetic information, the genomic era significantly reduced the resources needed to obtain large amount of relevant data (Rossetto & Henry 2014; Williams et al. 2014). Single-nucleotide polymorphism (SNP) data derived from next-generation sequencing technologies are particularly suitable for defining boundaries between lineages, as well as offering insights into adaptive processes (Funk et al. 2012; Allendorf 2017). A new era of evolutionarily informed restoration opportunities is arising, and it should be possible to combine new technologies within a sampling and interpretative framework that can provide the necessary data efficiently and cost effectively.

Within this context our objective is to establish a sampling and analytical protocol that enables the gathering of replicated multispecies evolutionary information (based on genomic, environmental, and distributional datasets). We introduce "Restore and Renew," a replicable framework that interprets uniformly gathered evolutionary, ecological, and genetic data across many species to meet the needs of restoration practitioners. The information is presented in an open access, easy to interpret, interactive webtool that facilitates the application of any provenance-based scenario for many species at any site of interest. The Restore and Renew webtool (www.restore-and-renew.org.au) is a continuously expandable system (in number of species and interpretative outputs) that supports restoration practices, and enables the establishment of experimental trials and monitoring procedures.

## Methods

### Sampling Strategy

Restore and Renew aims to equip restoration practitioners and land managers with a summary of pertinent evolutionary, environmental, and ecological information for a selected species and site (Rossetto 2017). The primary aim is to gather information on genetic provenance, ecology, and climatic suitability for over 100 Australian plant species commonly used in

restoration work. The target species were selected after consultation with restoration practitioners, nurseries, and ecologists to ensure even representation of commonly used taxa as well as broad functional, geographic, and ecological diversity (Hogbin & Rossetto 2014).

To achieve its objectives, Restore and Renew needed a sampling strategy that ensured even representation across the environmental and geographical distribution of each species while maximizing (when possible) between-species overlaps. To achieve maximum informative power while maintaining resource-effectiveness, we focused on maximizing the number of sites sampled across the distribution of each species (Prunier et al. 2013). Generally, and not surprisingly, sampling is logistically complex and the most resource-intensive component of the project. A brief description of the Restore and Renew framework follows:

(1) Distribution and environmental data is used to select between 30 and 50 representative sites that include geographic and environmental variation across the distribution of the target species. This is done simultaneously for multiple species to maximize site overlaps, thus simplifying collection logistics and enhancing multispecies analyses and interpretations.
(2) After subsampling in preliminary trials we settled on the collection of six individuals per species, per site to maintain efficiency in costs and logistics while providing sufficient genomic information at site and whole-species levels (Prunier et al. 2013; Lotterhos & Whitlock 2015). Herbarium vouchers are collected when no local voucher has previously been collected, if other projects require a voucher, or when there is taxonomic uncertainty.
(3) A dedicated application (App) built using iFormBuilder (Zerion Software, Inc.) is used on mobile devices to collect location and environmental data for every sample of every species in the field. These data include individual specimen barcodes, location information (e.g. latitude, longitude, and altitude recorded using the GPS on a mobile device), and other metadata relevant to the analytical approaches. We also collect additional ecological information that can help the interpretation of unexpected patterns. This information is used to inform further sampling, to support future fine-scale studies, and to maintain high standards of data capture for the National Herbarium of New South Wales (NSW) and other data repositories such as BioNet (http://www.bionet.nsw.gov.au/).
(4) Prebarcoded envelopes with an individual reference number (IRN) are used for collecting and storing tissue samples. Barcodes are scanned in the field using the App, and the IRN tracks the sample all the way through to the final data analysis.
(5) Samples are freeze-dried and stored in airtight containers with silica gel at the National Herbarium of NSW until they are processed.
(6) By mapping collection metadata, sampling progress for each species is monitored to assess species-specific progress and update field-collection plans to ensure that sampling covers the entire distribution of species as well as the breadth of environmental conditions they grow in.
(7) Once the full complement for a target species has been sampled, leaf tissue is placed in a randomized order in 96-well DNA extraction plates and sent to DArT Pty Ltd (www.diversityarrays.com), for DNA extraction and DArTseq genetic analysis.
(8) Genotype by sequencing (DArTseq) data are downloaded as .csv files and processed for quality control, analyses, and interpretation using our custom workflows.
(9) The analytical workflows (described in greater detail below) involve the delimitation of the area local genetic provenance through the use of general dissimilarity models (GDMs), and climate matching to a specified restoration site for current and future (2050) conditions.
(10) Species distribution data, genomic data, and climate data are then brought together and presented within the Restore and Renew webtool (www.restore-and-renew.org.au).

#### Genomic Data Analyses

We genotype the samples collected for each species using an approach developed by Diversity Arrays Technology Pty Ltd (DArT). DArT were the providers of choice because of logistic ease of access to their technology, and the need for continuing technical support in such a varied and complex project. DArT also provides DNA extraction and sequencing services, as well as data storage and handling, thus simplifying the infrastructural requirements of the Restore and Renew project. DArTseq is a high throughput, cost-effective restriction-based, reduced representation genome sequencing method that simultaneously assays thousands of markers, as well as codominant SNPs across the genome (Sansaloni et al. 2011; Kilian et al. 2012; Cruz et al. 2013). Reduced representation sequencing approaches, such as DArTseq, enable the cost-effective investigation of evolutionary processes at a genomic scale and the fine-scale examination of genetic variation across landscapes (Bragg et al. 2015). Recent studies using DArTseq have found that they are informative for understanding relationships among populations and species, particularly in closely related lineages (Yang et al. 2016; Melville et al. 2017; Rutherford et al. 2018).

The genome reduction and library construction method of DArTseq principally follows the methods described by Kilian et al. (2012) and Cruz et al. (2013). The genome complexity reduction performs a "methyl filtration" function achieved using at least one methylation-sensitive enzyme that does not cut DNA with a methylated cytosine base in the CXG motif. This feature enables a strict selection of hypomethylated sequences for the genomic libraries. In plant genomes, these sequences largely correspond to single/low copy sequences, mostly within (or closely linked to) genes, and distributed evenly across the genome in gene-rich areas (see Garavito et al. 2016 as an example of the distribution of DArTseq markers across the *Coffea* genome). The resulting sequencing library is largely depleted of paralog sequences allowing selection of true SNPs by the DArTsoft14 analytical pipeline.

Genotype data are generated using proprietary analytical pipelines (DArT). Briefly, this consists of processing fastq files to remove poor-quality sequences prior to marker calling. Identical sequences are collapsed and analyzed using a proprietary algorithm (DArT) to correct low-quality bases in a singleton tag by collapsing tags supported by multiple sequences as reference sequences. The corrected sequences are used in a secondary proprietary pipeline (DArT) which evokes SNP calling algorithms (DArTsoft14). These workflows and associated technical parameters have been validated by DArT Pty Ltd in controlled cross populations tests in which Mendelian behavior of markers can be rigorously tested. Approximately 30% of genotyped samples are reprocessed as technical replicates to measure the consistency of allele calls and calculate an index of reproducibility for each locus. Sequencing is performed to an average read depth of greater than 30 reads per locus. The DArTsoft14 analytical pipeline also includes a BLAST search of all loci to remove putative microbe contaminants. This step is essential when working with tissue collected from natural populations. Finally, all markers validated through the DArTsoft14 pipeline are stored in the DArTdb storage system, which allows for datasets processed at different times to be coanalyzed without the inclusion of biological replicates, thus simplifying multistep sampling and analyses when needed.

The genotype data are analyzed using a workflow (implemented in R; R Core Team 2016) that we specifically developed for Restore and Renew and consists of three main steps. We first calculate a range of descriptive statistics for the dataset, filter out SNP loci of poor quality (reproducibility average <0.96, genotypes missing in >20% of samples), and identify poor quality samples (samples missing data in a large proportion of loci). Second, we perform a series of preliminary population genetics analyses. The results are inspected to determine whether additional sampling is needed, or if there are "outlier" samples, as can occur due to errors, such as taxonomic misidentification, or real processes, such as recent hybridization. Outlier samples that are likely due to errors are removed, and this step is repeated to gain a preliminary understanding of patterns of genetic variation. Finally, we perform a set of analyses on a final dataset and build statistical models of genetic differentiation that are used in the Restore and Renew webtool. This begins with an analysis of population genetic stratification using sparse non-negative matrix factorization (sNMF) (Frichot et al. 2014) and R package landscape and ecological association (LEA) (Frichot & Fancois 2015), and the estimation of ancestry coefficients for each sample. This also includes principal component analysis (R package adegenet; Jombart & Ahmed 2011), and estimation of $F_{ST}$ values between all pairs of sampling sites (R package SNPRelate; Zheng et al. 2012).

#### Generalized Dissimilarity Models

Once the preliminary steps are completed, we estimate the models that are used to generate the information provided to practitioners (R package gdm; Manion et al. 2017). These are GDMs (Ferrier et al. 2007), which describe the level of genetic dissimilarity (differentiation) between sites as a function of spatial distance, as well as other predictive variables (Shryock et al. 2017; Supple et al. 2018). For species where spatial distance provides a good description of differentiation between population pairs, the GDM simply describes this dependence of population pairwise $F_{ST}$ on spatial distance. We use this model to predict the $F_{ST}$ between a pair of points based on their distance apart, or how far from a focal point—e.g. a site that is being revegetated—a threshold value of $F_{ST}$ is likely to be reached. For other species, genetic differentiation between populations will be affected by environmental variables (isolation by environment) or by population stratification arising from historical demographic processes including divergence and evolution in isolation (potentially followed by secondary contact). Here, we can use GDM models that have environmental variables (Supple et al. 2018) or ancestry coefficients (Shryock et al. 2017) as covariates, along with spatial distance.

Having estimated a GDM and used it to make predictions for $F_{ST}$ across the landscape, we then make a prediction for the genetically local area surrounding a specific location by choosing a threshold value of differentiation (in this case, $F_{ST} = 0.3$). The prediction of the genetically local area for a nominated location can be generated and displayed rapidly. We note that in the future, the availability of relevant data from an increasing amount of species might help us to further refine several aspects of the workflow. First, to generate a local area, it is necessary to select a threshold for differentiation. We chose the value of $F_{ST} = 0.3$ empirically, by examining outcomes for different species and testing different threshold values. If needed, it would be simple to adjust this threshold in the future for all or some species. For instance, we could reduce this value for species presenting increased risk of outbreeding depression, or if a more appropriate value was identified via experimentation. Second, it is likely we will encounter species with patterns of genetic variation that are not amenable to characterization with GDM. This could occur where a species has expanded across the landscape very recently, or where unobserved drivers have led to heterogeneity in the levels of differentiation among demes in different regions. GDM would be expected to fit such data poorly, resulting in a small percentage of null deviance explained. Here our advice to practitioners will vary depending on the magnitudes of observed levels of differentiation. For instance, if differentiation is hard to characterize with a spatial model, but is always very small, we will indicate that broad collections are unlikely to result in mixing of highly differentiated material. On the other hand, if differentiation is spatially heterogeneous and can reach high levels across small distances, we will indicate that collecting material from spatially separated sites carries a risk of mixing strongly differentiated material.

We continue to explore different methods of measuring and characterizing differentiation, including more explicit consideration of variation among individuals within our collection sites.

#### Species Distribution Data, Environmental Niche Modeling, and Climate Matching

Occurrence data for the targeted Restore and Renew species is obtained from the Atlas of Living Australia (www.ala.org

.au) and filtered to include only herbarium records, and to exclude erroneous records. Current climate is derived from monthly minimum and maximum temperatures and rainfall as 30-year averages for the period 1983–2012 using gridded climate data from the eMAST repository (http://dap.nci.org.au/thredds/catalog.html). Future climate conditions are derived from a 30-year climate average using data from global circulation models and Earth system models produced for the Intergovernmental Panel on Climate Change Fifth Assessment Report (Pachauri et al. 2014). We use an ensemble of seven climate models with demonstrated skill in modeling climate in the Australia region (CSIRO and Bureau of Meteorology 2015), and we select model runs driven by a moderate climate change scenario (RCP 4.5). The 30-year climatology we chose is for the near future and is centered on 2050. These data are downscaled onto the same grid as the eMAST data. Details of the downscaling method, and justification for selection of climate models and RCP 4.5, are provided in Appendix S1, Supporting Information.

To provide guidance on climate-matching and future-proofing options, we implement a simple box-car matching method similar to the technique used in the original Bioclim technique (Busby 1991) to define a spatial region around each user-selected location with climate conditions broadly similar to those at the location. Other methods have been devised for climatic matching (Robertson et al. 2001; Broadmeadow et al. 2005; Robertson et al. 2008; Grenier et al. 2012) but they all require extensive computations to implement, which is an impediment to the interactive nature of a web-based decision-support tool. We restrict the climate variables to a representative temperature (mean annual temperature [MAT]) and precipitation (mean annual precipitation [MAP]). A challenge for all climate-matching methods is selecting a threshold or cut-off to provide a binary spatial-layer indicating match versus no-match states. Ideally, any such decision rule should be based on knowledge of the fundamental niche of the organism, but this is not available in most contexts. We therefore select heuristic thresholds for temperature and precipitation of 1.5°C and 300 mm, respectively.

Grid cells matching current climate at the user-selected location are identified as those simultaneously falling within MAT ± 1.5°C and MAP ± 300 mm where the MAT and MAP values represent the current or observed climate conditions at the location. Grid cells matching predicted future climate at the location are computed in the same way with current MAT and MAP replaced by future values. The future climate layer represents climate conditions present in the landscape *now* which are predicted to occur at the user-selected restoration site in 2050. It allows restoration practitioners to consider seed collections from populations of the target species which may be pre-adapted to conditions expected at the restoration site. The webtool asks practitioners to use direct observation of local soil attributes to further refine search areas for seed collections.

We also produced ENMs using the machine-learning method MaxEnt (Phillips et al. 2006). Predictor variables used included bioclimatic variables (derived from the previously described climate data), soil composition variables (percent sand, silt, and clay), and the topographic variables, Topographic Wetness Index, and Topographic Position Index. Soil and topographic data were obtained from the CSIRO data portal (https://data.csiro.au/dap, last accessed 30 August, 2018). Details of the model-fitting procedure are provided in Appendix S1. MaxEnt models fitted to current climate were projected onto 2050 conditions with soil and topographic variables considered unchanged over the time span. The role of the fitted niche models was to provide an indication of overall environmental suitability and the nature of changes expected for the species across its distribution in NSW. ENMs are not suitable for climate matching for reasons explained in Appendix S1. To assist web tool users to understand the nature of projected changes in environmental suitability, we included a bar graph showing changes of environmental suitability across five suitability classes ranging from very low to very high suitability (details of this graphical tool are included in Appendix S1).

#### Illustrative Examples

The initial scope of Restore and Renew focuses on the State of NSW, Australia. NSW is a large state (around 800,000 $km^2$) characterized by the great dividing range running from north to south and separating the wet coastal vegetation from the drier inland. The diversity of terrestrial ecosystems leads to a range of complex vegetation types (comprising over 5,000 native vascular plant species) that include semi-arid shrublands, temperate grasslands, rainforests, eucalypt-dominated forests, chenopod shrublands, heathlands, as well as alpine and subalpine vegetation. About 40% of the state has been cleared of native vegetation, and overall vegetation condition is considered good in only 9% of NSW, with a similar amount of land coming under some type of protection agreement (https://soe.environment.gov.au/).

Restore and Renew is currently targeting over 100 species commonly used in restoration across NSW. To illustrate how the approach and workflows developed can respond to different evolutionary histories and dynamics, we present relevant models and predictions for two species with contrasting distributions and patterns of genetic variation: *Westringia fruticosa* (Lamiaceae; Fig. 1A) and *Acacia suaveolens* (Fabaceae; Fig. 1C).

*Westringia fruticosa* (Willd.) Druce (native rosemary) is a dense spreading shrub up to 1.5 m high. It has gray-green, narrow lanceolate leaves and the flowers are white, mostly found in the upper leaf axils throughout the year. Although little is known about the reproductive biology and ecology of the species targeted by this project, flower structure suggests that *W. fruticosa* is an outcrosser and most likely insect pollinated. Little is known about seed dispersal. Its natural distribution is restricted to the coastal heath near the ocean and along harbor foreshores of NSW. It is commonly used as a garden plant and some of the cultivars are hybrids and these should be categorically avoided in ecological restoration work (and were avoided in the sampling strategy). This species was selected to illustrate an example of patterns of genetic variation across the landscape that can be modeled quite effectively using spatial distance only (isolation by distance [IBD]) and without covariates such as environmental variables or ancestry coefficients (Fig. 1B). Here,

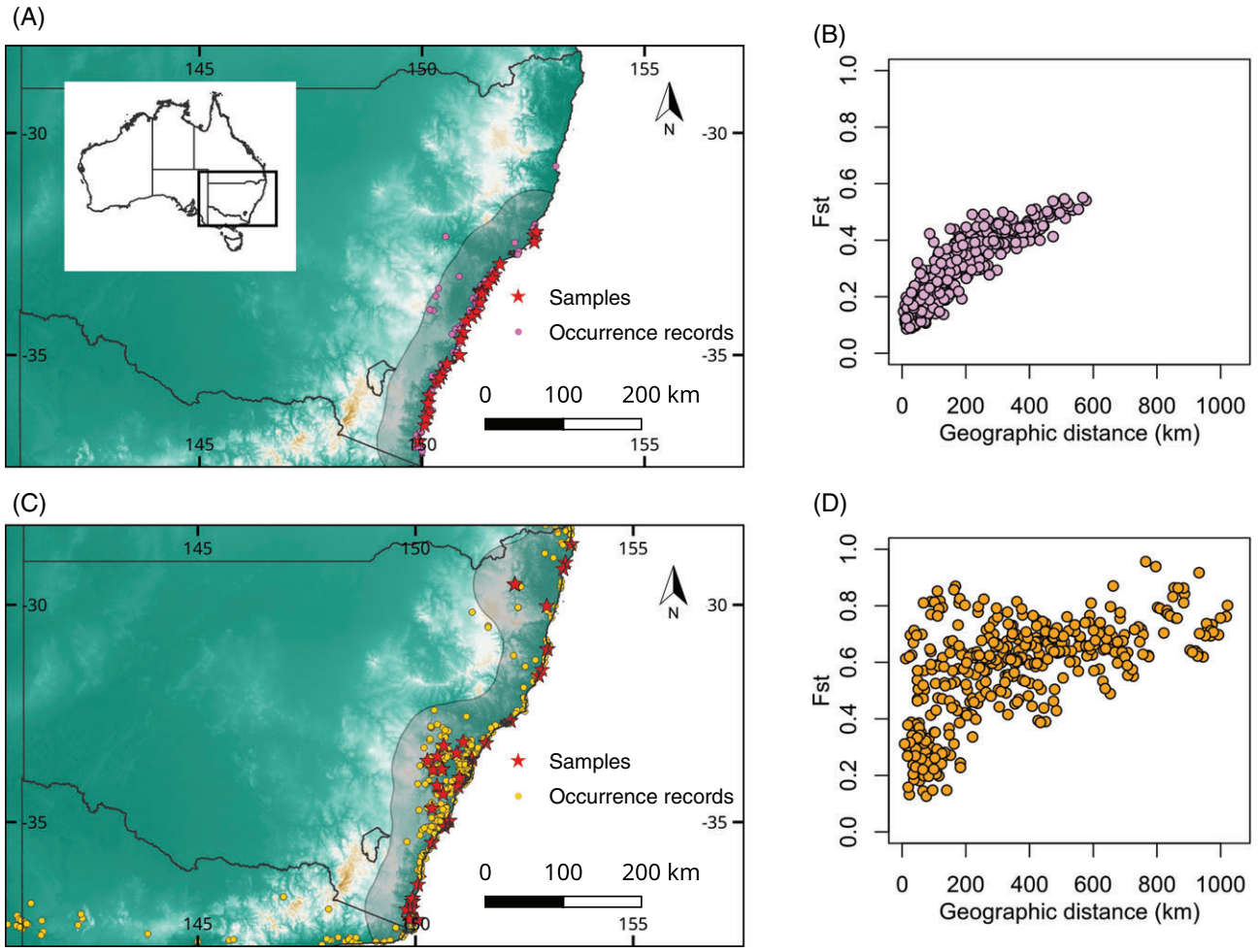



Figure 1. Plots illustrating the geographical ranges and distribution of genetic differentiation across space for exemplar species *Westringia fruticosa* (A, B) and *Acacia suaveolens* (C, D). For each, a map of occurrence records and sampling locations is shown (A, C), and pairwise genetic differentiation between populations ($F_{ST}$) is plotted as a function of spatial distance (B, D). Shaded regions represent the confidence envelope for each fitted generalized dissimilarity model.

we present data based on 155 samples collected at 27 sites across the range.

*Acacia suaveolens* (Sm.) Willd. (sweet wattle) is widely distributed along the east coast of Australia from Tasmania in the south to Queensland in the north. It is a prostrate-to-erect shrub to 2.5 m, and like many other acacias has a short life span, rarely living longer than 15 years and flowering early, usually within the second year after germination. It occurs in a range of soils and is tolerant of low temperatures. *A. suaveolens* is likely to be open pollinated, seed fall is passive and thought to be followed by short distance dispersal by ants, with recruitment generally occurring after fire (Auld 1986). This early colonizing shrub was selected to illustrate the case in which spatial distance alone is insufficient to explain genetic variation across the landscape, and where we need to build a model incorporating other explanatory variables (Fig. 1D). The dataset was finalized after excluding samples that were likely misidentified, or that had a large proportion of missing genotypes. Here, we present data for 166 samples collected at 29 sites across the range.

## Results

In *Westringia fruticosa* we observed patterns of genetic variation across the landscape that were consistent with IBD (Fig. 1B). When we summarized genetic variation among samples with a principal components analysis, we observed that the first principal component was correlated with latitude. Also, there was a monotonic increase in pairwise differentiation ($F_{ST}$) between populations as a function of their geographic distance. We fit a GDM to these pairwise differentiation data and explained 81% of null deviance in pairwise $F_{ST}$ using spatial distance as the predictor variable. We used this model to predict the $F_{ST}$ between a "focal" point in space—a nominal site for restoration—and all other grid cells (Fig. 2A). If we mask all cells predicted to be differentiated, from the focal point with $F_{ST}$ greater than the threshold value, we have produced a prediction for the genetically "local" area (Fig. 2B).

In contrast, pairwise spatial distances between *A. suaveolens* populations were not sufficient to explain levels of genetic

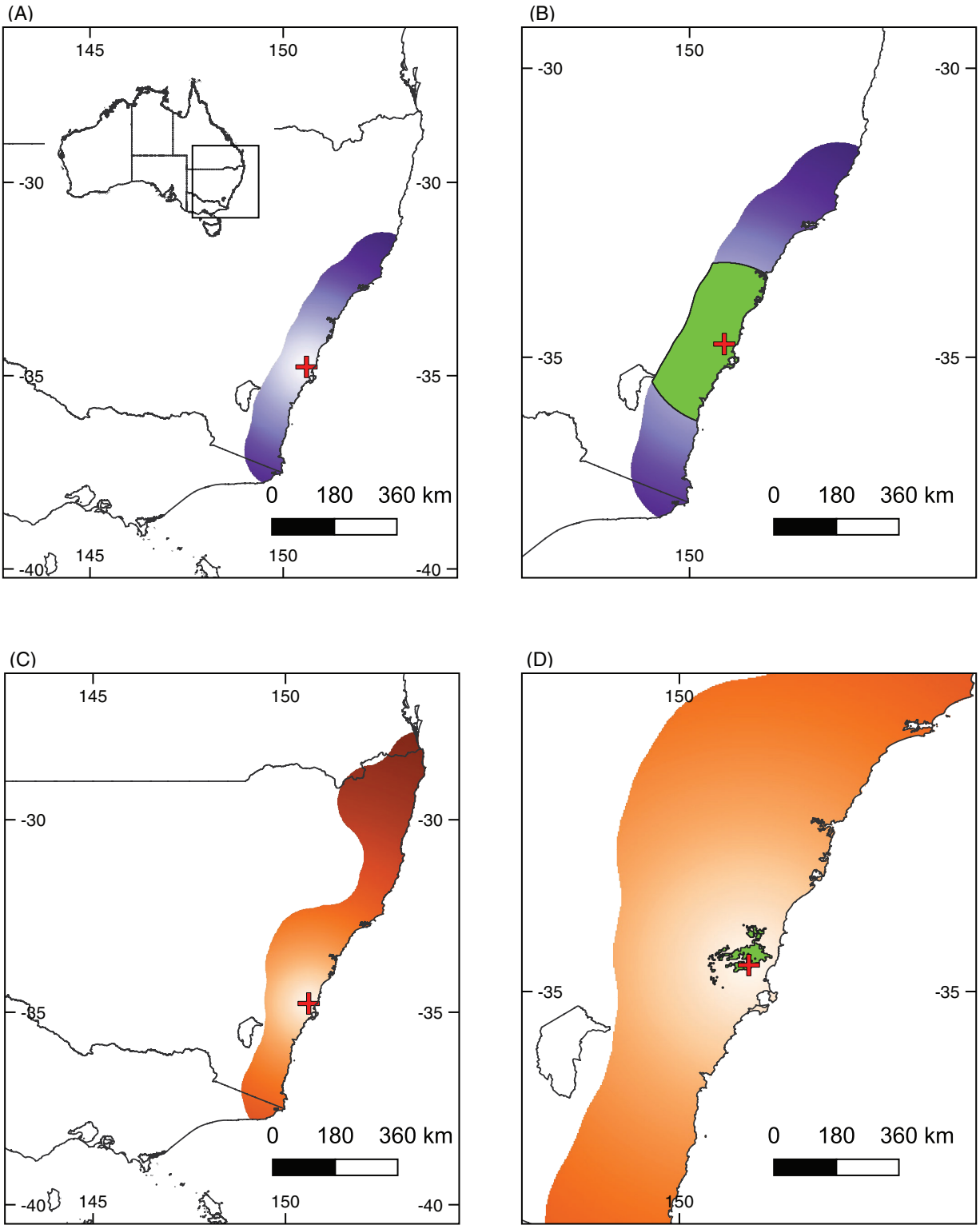
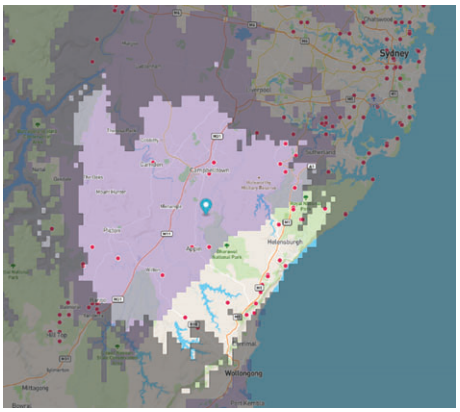



Figure 2. Predicted levels of genetic differentiation from a selected location (shown by a cross). For *Westringia fruticosa* (A, B) and *Acacia suaveolens* (C, D), a generalized dissimilarity model was fit to pairwise population genetic differentiation. For each species, the corresponding model was used to predict levels of differentiation (less intense shading, higher differentiation) for all other locations across the landscape (within a buffered hull around sampled locations) (A, C). It is then possible to set a threshold level of differentiation, exclude points with values greater than the threshold, and identify an area of genetically local provenance (green areas; B, D). Boundaries can be uniform when responding to simple isolation by distance models (*W. fruticosa*) or convoluted when following more complex criteria (such altitude in the case of *A. suaveolens*).

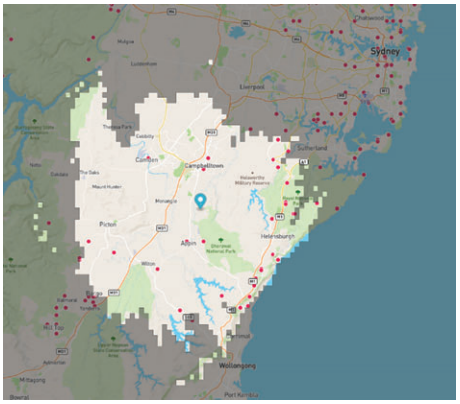
**Box 1. Interpreting the Restore and Renew data to replicate ecological restoration scenarios.**

A simplified illustrative example for *Acacia suaveolens* is presented, to represent how the information obtained through the Restore and Renew webtool could be used to replicate four restoration scenarios. For *Acacia suaveolens*, the analytical processes used to identify the geographical range of local genetic provenance at the selected site take into consideration elevation and spatial distance. As a result, in regions where abrupt changes in elevation occur within a small geographical area, the area identified as the local provenance (i.e. nongrayed in the maps) has irregular boundaries and/or some fragmentation. This is a representation of how, based on empirical data, the model responds to the presence of steep local altitudinal gradients within the distribution of this species.

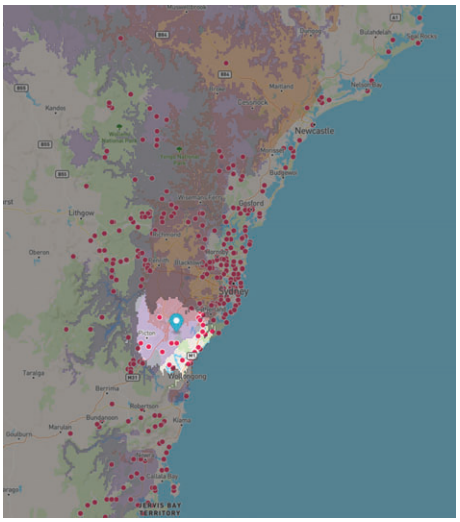
Local provenancing (McKay et al. 2005): with the information modeled and presented, seed sources can be selected from local climatic and environmental conditions (i.e. similar to those within the site to be restored) within the identified genetic neighborhood. This would involve sampling from multiple locations in the areas that fall within the overlap between the local genetic provenance (nongrayed area on the map), and the climate-matched regions (highlighted in purple). In many cases the area might also be sufficiently large to enable the maximization of genetic diversity even within this relatively narrow-scoped strategy.




Composite provenancing (Broadhurst et al. 2008): following the local boundaries of a genetic neighborhood, seeds can be sourced from a range of sites with varying environmental and climatic conditions. These include site-matched, future-matched, and unmatched (any area within the nongrayed zone of the map). Small proportion of material can also be obtained from sites outside that neighborhood (i.e. the grayed area of the map).

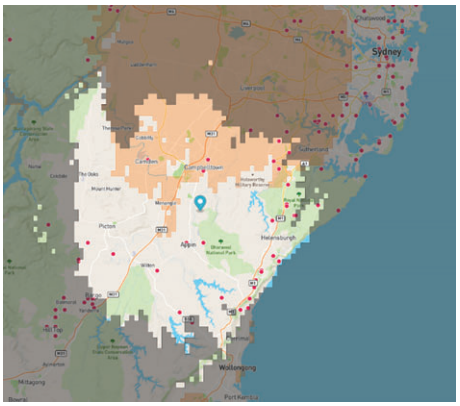



Climate-adjusted provenancing (Prober et al. 2015): knowing local neighborhood boundaries will enable sampling seed sources from local genetic and local environmental conditions. Additional sources can be selected outside the local genetic neighborhood that are incrementally similar to the climatic conditions predicted based on future climate change scenarios presented (and highlighted in orange in this and in the following image).



**Box 1. Continued**

Predictive provenancing (Sgro et al. 2011): modeled surfaces of future climate enable the selection of seed sources from areas that are already experiencing the climatic conditions predicted based on future climate change scenarios. In some circumstances of course, these can include sites that are within the local genetic neighborhood.




differentiation (Appendix S2). Some population pairs that are close together in space had large values of genetic differentiation. Also, in a principal components analysis, the first principal component is associated with elevation (Appendix S2). We do not suggest here that elevation directly influences patterns of genetic variation, but is instead acting as a useful surrogate, perhaps helping to account for variation between lineages that tend to occur at different elevations. This is corroborated by an sNMF analysis, which identifies different ancestral populations inhabiting a narrow strip along the coast and nearby upland areas.

Based on these observations, for *A. suaveolens* we fit a GDM that had spatial distance and elevation as explanatory variables. This model explained 48% of null deviance in pairwise $F_{ST}$. When we used this model to predict $F_{ST}$ from a focal site, $F_{ST}$ accumulated to larger values over much shorter distances relative to *W. fruticosa*, as expected. Correspondingly, the predicted genetically local areas tended to be substantially smaller for *A. suaveolens*.

In summary, for two contrasting test cases, a GDM explains patterns of genetic differentiation across the landscape and provides predictions for the size and shape of the local genetic neighborhood at any selected point (that falls within a buffered hull around our sampling locations). For *W. fruticosa,* which largely conforms to an IBD model, the local neighborhood is a circle with the radius determined by the prevailing rate at which differentiation accumulates with distance across the landscape (Fig. 2B). For *Acacia suaveolens*, differentiation is associated strongly with differences in elevation and spatial distance, and therefore the local neighborhood boundaries are not evenly circular and are narrower where the area crosses a strong elevation gradient. Box 1 illustrates how the *A. suaveolens* case study is presented and interpreted within the Restore and Renew webtool. Genetic provenance boundaries are calculated on the fly relative to the site/species combination selected by the user, and presented on the map as a highlighted area. Climate matching found at the selected site is highlighted, and the user has the option of also highlighting areas matching the climate that will occur at that specified site in 2050. Details about the webtool and how it can be used are available via the "help" button at www.restore-and-renew.org.au.

## Discussion

There is a growing need, in ecological restoration and revegetation practices, for genetic information in support of collection strategies that aspire to the survival and persistence of populations (Mijangos et al. 2015). Here we describe a framework to collect, interpret, and communicate multispecies provenance information to restoration practitioners efficiently (in speed of computation) and effectively (in ease of visualization and interpretation). Restore and Renew will continue to add new species, as well as layers of information and interpretative support, as new pipelines for rapid data mining are developed.

Since species are different, illustrated by the two contrasting cases above, the objective of Restore and Renew is to provide easy-to-access information for practitioners to support empirically informed choices that are relevant to their needs and opportunities. The data are presented within the webtool so that genetic differentiation boundaries can be interpreted within current or future climatic contexts to fine-tune planting strategies for the site/species combination selected. The availability of such boundaries provides new opportunity to trial restoration strategies, and hopefully supports the design of experiments testing the validity of contrasting approaches.

As data for more species are obtained, we expect to benefit from a broader understanding of interspecific variation in landscape genetic patterns. This will bolster our ability to make inferences about processes driving differentiation on the landscape, and lead to technical refinements in our workflows that characterize these patterns of differentiation. For instance, as we gain a better understanding of the range of variation among species, especially in relation to trait and mating system variation (which are currently mostly unknown), we will refine our approach to modeling and delimiting the genetically local area.

Restore and Renew exploits recent technical developments to establish a large-scale program of data collection and analyses

with broad applications and implications for restoration, revegetation, and biodiversity management. For many species, it provides an empirically based understanding of how genetic diversity is distributed across the landscape, and within current and future climate expectations. This represents a first critical step towards supporting restoration practitioners in their choice of material and strategy, and in the development of experimental and monitoring plans. Once combined with practical considerations related to seed collection, preparation, and storage, as well as site preparation, planting, and monitoring (Guja et al. 2015) this information can provide a considerable improvement of restoration outcomes in general.

## Acknowledgments

This work was supported by funding from the Royal Botanic Gardens and Domain Trust, the NSW Environmental Trust, and HSBC Bank Australia (Principal Conservation Partner). The authors thank Diversity Arrays Technology Pty Ltd for providing technical support. The authors thank Joel Cohen, Monica Fahey, Susan Rutherford, Sam Yap, and our team of volunteers for field sampling and curation.

## LITERATURE CITED

Allendorf FW (2017) Genetics and the conservation of natural populations: allozymes to genomes. Molecular Ecology 26:420–430

Auld TD (1986) Population dynamics of the shrub *Acacia suaveolens* (Sm.) Willd.: dispersal and the dynamics of the soil seed-bank. Australian Journal of Ecology 11:235–254

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. New Phytologist 207:953–967

Breed M, Stead M, Ottewell K, Gardner M, Lowe A (2013) Which provenance and where? Seed sourcing strategies for revegetation in a changing environment. Conservation Genetics 14:1–10

Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, Yates C (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. Evolutionary Applications 1:587–597

Broadhurst L, Breed M, Lowe A, Bragg J, Catullo R, Coates D, et al. (2017) Genetic diversity and structure of the Australian flora. Diversity and Distributions 23:41–52

Broadmeadow MSJ, Ray D, Samuel CJA (2005) Climate change and the future for broadleaved tree species in Britain. Forestry 78:145–161

Busby JR (1991) BIOCLIM—a bioclimate analysis and prediction system. Plant Protection Quarterly 6:8–9

Crowe KA, Parker WH (2008) Using portfolio theory to guide reforestation and restoration under climate change scenarios. Climatic Change 89:355–370

Cruz VMV, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic diversity assessment of the US collection of the new oilseed crop *lesquerella* and related species. PLoS One 8:e64062

CSIRO and Bureau of Meteorology (2015) Climate change in Australia. Information for Australia's natural resource management regions. Technical report. CSIRO and Bureau of Meteorology, Melbourne, Australia

Falk DA, Palmer MA, Zedler JB (eds) (2006) Foundations of restoration ecology. Island Press, Washington D.C.

Fenster CB, Galloway LF (2000) Inbreeding and outbreeding depression in natural populations of *Chamaecrista fasciculata* (Fabaceae). Conservation Biology 14:1406–1412

Ferrier S, Manion G, Elith J, Richardson K (2007) Using generalized dissimilarity modeling to analyze and predict patters of beta diversity in regional biodiversity assessment. Diversity and Distributions 13:252–264

Frankham R, Ballou JD, Eldridge MD, Lacy RC, Ralls K, Dudash MR, Fenster CB (2011) Predicting the probability of outbreeding depression. Conservation Biology 25:465–475

Frankham R, Ballou JD, Ralls K, Eldridge MD, Dudash MR, Fenster CB, Lacy RC, Sunnucks P (2017) Genetic management of fragmented animal and plant populations. Oxford University Press, London, United Kingom

Frichot E, Fancois O (2015) LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution 6:925–929

Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends in Ecology and Evolution 27:489–496

Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. BMC Plant Biology 16:242

Grenier P, Parent A-C, Huard D, Anctil F, Chaumont D (2012) An assessment of six dissimilarity metrics for climate analogs. Journal of Applied Meteorology and Climatology 52:733–752

Guja LK, Broadhurst LM, Brown AH, Bush D, Cochrane A, Merritt DJ, Wood J (2015) Genetic diversity is a significant but not the only consideration for effective ex situ plant conservation: response to Hoban and Schlarbaum. Biological Conservation 184:467–468

Hancock N, Hughes L (2012) How far is it to your local? A survey on local provenance use in New South Wales. Ecological Management and Restoration 13:259–266

Hedrick PW, Garcia-Dorado A (2016) Understanding inbreeding depression, purging, and genetic rescue. Trends in Ecology and Evolution 31:940–952

Hogbin T, Rossetto M (2014) Restore and Renew NSW: online survey of NSW restoration practitioners identifies species considered most useful in restoration. Australasian Plant Conservation 22:26

Jombart T, Ahmed I (2011) adegenet 13-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3071–3071

Kettenring KM, Mercer KL, Reinhardt Adams C, Hines J (2014) Application of genetic diversity–ecosystem function research to ecological restoration. Journal of Applied Ecology 51:339–348

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. Data Production and Analysis in Population Genomics 888:67–89

Krauss SL, Koch JM (2004) Methodological insights: rapid genetic delineation of provenance for plant community restoration. Journal of Applied Ecology 41:1162–1173

Lesica P, Allendorf FW (1999) Ecological genetics and the restoration of plant communities: mix or match? Restoration Ecology 7:42–50

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology 24:1031–1046

Manion G, Lisk M, Ferrier S, Nieto-Lugilde D, Mokany K, Fitzpatrick MC (2017) gdm: generalized dissimilarity modeling. R package version 132

McDonald T, Jonson J, Dixon KW (2016) National standards for the practice of ecological restoration in Australia. Restoration Ecology 24(S1):S4–S32

McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?"—a review of practical and conceptual issues in the genetics of restoration. Restoration Ecology 13:432–440

Melville J, Haines ML, Boysen K, Hodkinson L, Kilian A, Smith Date KL, Parris KM (2017) Identifying hybridization and admixture using SNPs: application of the DArTseq platform in phylogeographic research on vertebrates. Royal Society Open Science 4:161061

Mijangos JL, Pacioni C, Spencer P, Craig MD (2015) Contribution of genetics to ecological restoration. Molecular Ecology 24:22–37

Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, Church JA, Clarke L, Dahe Q, Dasgupta P, Dubash NK (2014) Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Perring MP, Standish RJ, Price JN, Craig MD, Erickson TE, Ruthrof KX, Whiteley AS, Valentine LE, Hobbs RJ (2015) Advances in restoration ecology: rising to the challenges of the coming decades. Ecosphere 6:131

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modeling 190:231–259

Prober SM, Byrne M, McLean EH, Steane DA, Potts BM, Vaillancourt RE, Stock WD (2015) Climate-adjusted provenancing: a strategy for climate-resilient ecological restoration. Frontiers in Ecology and Evolution 3:65

Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P, Lena JP (2013) Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. Molecular Ecology 22:5516–5530

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rice KJ, Emery NC (2003) Managing microevolution: restoration in the face of global change. Frontiers in Ecology and the Environment 1:469–478

Robertson MP, Caithness N, Villet MH (2001) A PCA-based modeling technique for predicting environmental suitability for organisms from presence records. Diversity and Distributions 7:15–27

Robertson MP, Kriticos DJ, Zachariades C (2008) Climate matching techniques to narrow the search for biological control agents. Biological Control 46:442–452

Rossetto M (2017) Restore and Renew, a wealth of new information in support of ecological restoration practices. Australasian Plant Conservation 25:5–6

Rossetto M, Henry RJ (2014) Escape from the laboratory: new horizons for genetics. Trends in Plant Science 19:554–555

Rossetto M, Kooyman RM (2011) Predicting distributions by analyzing evolutionary, functional and environmental information—a review of population genetic studies on the Australian rainforest flora. Pages 305–322. In: Daniels JA (ed) Advances in Environmental Research. Vol 9. Nova Science Publishers, Hauppauge, New York

Rutherford S, Rossetto M, Bragg JG, McPherson H, Benson D, Bonser SP, Wilson PG (2018) Speciation in the presence of gene flow: population genomics of closely related and diverging eucalyptus species. Heredity 121:126–141

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of eucalyptus. BMC Proceedings 5:54

Sgro CM, Lowe AJ, Hoffmann AA (2011) Building evolutionary resilience for conserving biodiversity under climate change. Evolutionary Applications 4:326–337

Shryock DF, Havrilla CA, DeFalco LA, Esque TC, Custer NA, Wood TE (2017) Landscape genetic approaches to guide native plant restoration in the Mojave desert. Ecological Applications 27:429–445

Supple MA, Bragg JG, Broadhurst LM, Nicotra AB, Byrne M, Andrew RL, Widdup A, Aitken NC, Borevitz JO (2018) Landscape genomics prediction for restoration of a eucalyptus foundation species under climate change. eLife 7:e31835

Szűcs M, Melbourne BA, Tuff T, Weiss-Lehman C, Hufbauer RA (2017) Genetic and demographic founder effects have long-term fitness consequences for colonizing populations. Ecology Letters 20:436–444

Weeks AR, Sgro CM, Young AG, Frankham R, Mitchell NJ, Miller KA, et al. (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. Evolutionary Applications 4:709–725

Williams AV, Nevill PG, Krauss SL (2014) Next generation restoration genetics: applications and opportunities. Trends in Plant Science 19:529–537

Yang X, Ren R, Ray R, Xu J, Li P, Zhang M, Liu G, Yao X, Kilian A (2016) Genetic diversity and population structure of core watermelon (*Citrullus lanatus*) genotypes using DArTseq-based SNPs. Plant Genetic Resources 14:226–233

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328

## Supporting Information

The following information may be found in the online version of this article:

**File S1.** Supporting information.
**Appendix S1.** Climate data, downscaling and environmental niche modeling.
**Figure S1.** *Acacia suaveolens* (A) current ENM output map, (B) predicted 2050 ENM output map, (C) environmental change bar plot.
**Figure S2.** *Westringia fruticosa* (A) current ENM output map, (B) predicted 2050 ENM output map, (C) environmental change bar plot.
**Appendix S2.** Supplementary analytical details for an illustrative example—*Acacia suaveolens*.
**Figure S3.** Mean ancestry coefficients for the individuals of *Acacia suaveolens* at each sampled site, as inferred using sNMF.
**Figure S4.** Principal components analysis of genetic data, in relation to elevation.

*Coordinating Editor: Fernando Valladares*

*Received: 6 September, 2018; First decision: 1 October, 2018; Revised: 6 October, 2018; Accepted: 7 October, 2018; First published online: 25 November, 2018*