Review

CellPress

# Next generation restoration genetics: applications and opportunities

**Anna V. Williams[1,2], Paul G. Nevill[1,2], and Siegfried L. Krauss[1,2]**

[1] School of Plant Biology, The University of Western Australia, Crawley, WA 6009, Australia
[2] Kings Park and Botanic Garden, Botanic Gardens and Parks Authority, West Perth, WA 6005, Australia

**Restoration ecology is a young scientific discipline underpinning improvements in the rapid global expansion of ecological restoration. The application of molecular tools over the past 20 years has made an important contribution to understanding genetic factors influencing ecological restoration success. Here we illustrate how recent advances in next generation sequencing (NGS) methods are revolutionising the practical contribution of genetics to restoration. Novel applications include a dramatically enhanced capacity to measure adaptive variation for optimal seed sourcing, high-throughput assessment and monitoring of natural and restored biological communities aboveground and belowground, and gene expression analysis as a measure of genetic resilience of restored populations. Challenges remain in data generation, handling and analysis, and how best to apply NGS for practical outcomes in restoration.**

## Ecological restoration

Ecological restoration is the process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed [1]. The scale and demand for ecological restoration is large and rapidly growing – currently estimated to be a US$2 trillion/year activity worldwide [2], with significant targets such as the restoration of 150 million ha of disturbed and degraded land globally by 2020 [3]. Since the development of restoration ecology as a strong scientific discipline over the past 20 years [4], the fundamental importance of what can be termed 'genetic issues' has been increasingly recognised. A search within the journal *Restoration Ecology* (the key journal linking researchers and restoration practitioners, published by the Society for Ecological Restoration) with the term 'genetics' returned 143 papers published in the 21 years since the journal's inception in 1993. A similar search with the term 'molecular' returned 86 hits, emphasising the important contribution molecular markers (see Glossary) have made to restoration ecology. The past two decades have seen a remarkable development in the molecular toolkit available for restoration genetics, beginning with methods such as allozymes, which predate the polymerase chain reaction (PCR) revolution in the 1990s. With the development of PCR, measuring genetic variation at the DNA level became possible, principally with amplified fragment length polymorphism (AFLP) [5], microsatellites [6], and more recently through single nucleotide polymorphisms (SNPs) [7].

We are now experiencing another revolution with the development of next generation sequencing (NGS) technology. NGS describes a suite of new DNA sequencing technologies, the power of which lies in an ability to produce massive amounts of DNA sequence data at a low cost per

## Glossary

**Epigenetics:** changes in gene expression caused by mechanisms other than changes in the underlying DNA sequence.
**Expressed sequence tag:** a short fragment of DNA produced by reverse transcription of mRNA into DNA.
**Genome scan:** assessment and comparison of variation and behaviour at many genomic markers. A genome scan can differentiate neutral markers from outlier markers that presumably reflect non-neutral processes such as selection.
**Genotyping by sequencing:** approaches that allow a targeted fraction of the genome (a reduced representation library) to be sequenced with NGS rather than the entire genome, even in species with little or no previous genomic information and large genomes. The subset of the genome to be sequenced using these approaches may be targeted using restriction enzymes or capture probes or by sequencing the transcriptome.
**Metabarcoding:** a rapid method of biodiversity assessment that combines DNA taxonomy and NGS.
**Metagenomics:** the study of the collective genomic material contained in an environmental sample of microorganisms, facilitated by NGS of heterogeneous samples.
**Microsatellite:** short tandem repeat sequence, usually comprising a variable number of repeats of two to eight nucleotides. Different numbers of repeats result in different lengths of alleles.
**Molecular marker:** a heritable, measurable fragment of DNA that contributes data on patterns of genetic variation and its spatial structure within and among populations and individuals, and the processes affecting that pattern (e.g., selection, paternity, etc.).
**Next generation sequencing (NGS):** highly parallel DNA sequencing where thousands or millions of reads (sequences) are produced in one run. Current methods include single molecular real-time sequencing (Pacific Bio), pyrosequencing (454), sequencing by synthesis (Illumina), and sequencing by ligation (SOLiD). Methods in development include nanopore DNA sequencing and microscopy based techniques such as atomic force microscopy.
**Neutral marker:** genome locations that are not evolving directly in response to selection.
**Outlier marker:** genome locations that show patterns of variation that are extremely divergent from the rest of the genome.
**Population genomics:** genotyping hundreds to thousands of markers from across the genome in multiple divergent populations to identify outlier markers that have a high level of differentiation and are likely to be associated with genes involved in adaptation.
**Single nucleotide polymorphism (SNP):** nucleotide site in a DNA sequence where more than one nucleotide (G, A, T, or C) is present in the population.
**Transcriptome:** the set of all expressed RNA molecules produced in one or a population of cells, and represents the small percentage (ca. 5% in humans) of the genetic code transcribed into RNA molecules. The use of NGS to sequence and study transcriptomes at the nucleotide level is known as RNA-seq.

*Corresponding author:* Krauss, S.L. (skrauss@bgpa.wa.gov.au).
*Keywords:* restoration ecology; adaptation; population genomics; next generation sequencing; genotyping by sequencing; metabarcoding.

1360-1385/
 http://dx.doi.org/10.1016/j.tplants.2014.03.011

base, in a relatively short period of time. NGS is capable of producing billions of short read (50–800 bp) sequences in parallel, which makes possible the sequencing of whole genomes, or the genotyping by sequencing of hundreds of individuals at thousands of markers simultaneously, even in species for which little or no genetic information is currently available. Decreasing costs and reduced representation methods are making NGS available to non-model organisms [8,9], and even whole genome population genetic studies [10]. Detail on the concepts, platforms, challenges, and limitations can be found in many reviews [11–14].

In this review, we draw on the recent plant literature for examples that specifically address practical applications in the field of ecological restoration genetics, made possible only by harnessing the power of NGS. We begin by illustrating how the application of genetic markers to date has addressed the central issues in ecological restoration genetics.

## Ecological restoration genetics

Ecological restoration genetics is a discipline concerned with researching and understanding genetic issues that may impact on the practice of ecological restoration (Box 1) [15–17]. Underpinning these issues is the knowledge that almost all species show spatial genetic structure across their range, with varying levels of genetic variation within populations. Within a restoration context, genetic markers have been used to characterise the genetic variation within populations and the differentiation between them, with a primary focus on the issue of 'how local is local', where importance is placed on the use of local provenance seed so as to minimise negative impacts in restoration [17]. However, there is a robust debate on seed sourcing issues underway that is challenging the dogma of 'local is best'. Alternative strategies include composite provenancing, admixture provenancing, predictive sourcing for climate change, and novel systems (Figure 1) [17–24]. NGS offers powerful new opportunities to address these issues, as described below.

Restoration genetics has its roots in genecological research stemming back more than 200 years [25]. Common garden and reciprocal transplant provenance trials clearly demonstrate the important role of environmental selection in shaping genetic variation within species [26]. In a restoration context, this has been identified as a 'home site advantage' [27]. With the advent of isozyme markers in the 1950s, the ability to better quantify genetic variation within and among populations became possible [28]. This method has been applied to many hundreds of plant species, so that we have a good understanding of the relationships between life history traits and parameters of genetic variation and its spatial structure [29]. This understanding provides a powerful resource to infer genetic provenance guidelines from life history properties for species with no population genetic knowledge [30].

With the development of PCR, AFLP [5] and microsatellites [31] became the tools of choice for restoration genetics, enabling the assessment of genotypic variability across hundreds, even thousands, of markers and/or highly polymorphic loci. Practical applications include the delineation of local genetic provenance seed sourcing zones [32–38], the comparative assessment of diversity in restored and natural populations [39], the assessment of genetic distance between populations in outbreeding depression studies [40], the detection of genetic changes over generations during seed increase programs [41], for assessing the maintenance of genetic diversity in restored populations [42], for understanding the implications of fine scale spatial genetic structure for restoration plantings [43],

### Box 1. Key issues and questions in ecological restoration genetics

**Ecological restoration**: the process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed.

**Restoration ecology**: the science on which the practice of ecological restoration is based. Restoration ecology ideally provides clear concepts, models, tools, and methodologies for practitioners in support of their practice. Restoration ecologists also advance ecological theory by using restoration project sites as experimental areas.

**Provenance**: refers to the geographical origin of something; in an ecological restoration context, the source of propagules, typically seed. The extent, and biological significance, of the local provenance is a fundamental issue for successful ecological restoration (see Figure 1 in main text). Restoration practitioners ask: how local is local?

**Local adaptation**: whereby genotypes growing in their local environment display higher fitness than non-local genotypes introduced to the same environment – that is, locally adapted genotypes display a 'home site advantage', and introducing maladapted genotypes can impact on the success of ecological restoration. Key questions include: where are the potential source populations with genotypes that are best adapted to the restoration site? And, is it better to mix (provenances) than to match on highly disturbed sites?

**Inbreeding depression**: a reduction in fitness caused by breeding between genetically similar individuals, which may result from the too narrow sourcing of propagules for ecological restoration. Do fragmented potential source populations suffer from elevated inbreeding and therefore are their seeds compromising restoration success?

**Outbreeding depression**: a reduction in fitness caused by breeding between genetically dissimilar individuals, which may result from the too broad sourcing and mixing of propagules for ecological restoration. When, and at what spatial scale, does outbreeding depression impact on restoration success?

**Genetic swamping**: the displacement of local genotypes by non-indigenous genotypes with greater numbers and/or fitness. What is the impact on locally significant gene pools from the introduction of non-indigenous genotypes?

**Evolutionary potential**: whereby genetic diversity is representative of a restored populations potential to adapt to future environmental changes. Does admixture provenancing of seed sources increase the long-term success of restoration?

**Bottlenecks, founder events, and genetic drift**: whereby small population size and low genetic diversity in the founding population can result in a loss of genetic diversity in restored populations over time. How many founders are required to establish a genetically resilient restored population?

**Climate change**: whereby contemporary climate change is imposing strong selection pressures at a rapid rate, suggesting a need for predictive provenancing (see Figure 1 in main text). Can we identify source populations that are preadapted to future climate predicted for the restoration site?

**Restored population functionality**: the growing awareness that restoring functional ecosystem services is critical for self-sustaining, integrated, resilient restored meta-communities, and includes pollinator services, seed dispersal, food webs, invasion resistance, pest control, etc. Do pollinators follow plant establishment to facilitate reproductive functionality of a new plant community? Are the vectors of seed dispersal present?