

Sequencing of wild crop relatives to support the conservation and utilization of plant genetic resources

Robert J. Henry*

Queensland Alliance for Agriculture and Food Innovation, University of Queensland,
Brisbane, QLD 4072, Australia

Abstract

The re-sequencing of the genomes of wild crop relatives is a rapid method to determine the likely utility of the germplasm in crop improvement. The conservation of genetic resources both *in situ* and *ex situ* can be guided by information on the novelty of specific populations at the whole-genome and specific allele levels. The analysis of Australian wild relatives of rice, coffee, Macadamia and Eucalypts is being used to support crop improvement and enhance food and energy security. Rice populations that are novel sources of diversity in the A genome of rice have been characterized at the whole-genome level. This has demonstrated the novelty of these species and will support taxonomic revisions of the *Oryza* species. Variation in the genomes of plants from diverse environments defines strategies that might be employed to develop climate-resilient crop varieties. Eucalypt sequencing aims to support the selection of species and genotypes for use as new energy crops.

Keywords: whole-genome sequencing; wild crop relatives

Introduction

Advances in DNA sequencing (Edwards *et al.*, 2012) make whole-genome sequencing an attractive option for the characterization of plant genetic resources (Henry *et al.*, 2012). The sequencing of the genomes of plant genetic resources has great potential to support better management of both *in situ* and *ex situ* resources.

Examples of current applications of whole-genome sequencing to plant germplasm analysis in Australia include rice (*Oryza* species), coffee (*Coffea* sp.), Macadamia (*Macadamia* species) and Eucalypts (*Eucalyptus* and *Corymbia* species), Table 1.

Rice was domesticated in Asia, but it has close wild relatives in Australia. The analysis of the sequences of wild *Oryza* (Nock *et al.*, 2011) from Australia aims to define the uniqueness and resulting potential value

of the Australian populations as genetic resources for rice crop improvement (Waters *et al.*, 2012). This information is critical to the appropriate management of wild genetic resources for rice in Australia. Poorly characterized wild populations may be under threat, especially if rice production begins on a large scale in Northern Australia.

A redefinition of the coffee genus has included more species. *Coffea brassii* is a poorly known species found in Cape York in north Queensland. This species provides an opportunity to explore the unique characteristics of domesticated coffee by comparative whole-genome sequencing.

The progenitors of domesticated Macadamia are found in the subtropical rainforests of Australia near Brisbane, but the relatively recent domestication of Macadamia has taken place in Hawaii. The analysis of the sequences of wild *Macadamia* species will allow us to understand the relationships between the wild species in Australia and the domesticated varieties derived from Hawaii.

*Corresponding author. E-mail: robert.henry@uq.edu.au

Eucalypts cultivated worldwide are derived from material sourced in Australia in the last 200 years. Many recent plantings have continued to rely upon wild or close-to-wild germplasm. The analysis of the sequences of Eucalypts will assist us to understand their relationships and suitability for ongoing domestication for diverse uses, including potential new end-uses such as bioenergy and biomaterial production (Persil-Cetinkol *et al.*, 2012).

Experiences from these examples will guide further development of sequencing tools and strategies for the characterization and utilization of plant genetic resources. The sequencing of wild plants with special traits such as halophytes will provide additional information to support crop improvement.

Rice

The shotgun whole-genome sequencing of wild Australian rice varieties has confirmed the distinctness of the Australian populations. This result indicates that *Oryza rufipogon* in Australia may need to be recognized as one or more distinct species that represent a diverse source of germplasm for rice crop improvement. The analysis of the variation in genomes in relation to climate in wild populations provides a useful tool to identify strategies for breeding climate-resilient crops (Fitzgerald *et al.*, 2011; Shapter *et al.*, 2012). Whole-genome re-sequencing is being explored as a tool for application in rice crop breeding (Gopala Krishnan *et al.*, 2012), and sequencing of wild relatives will extend our knowledge of allelic diversity available for rice crop improvement. In this research, genome sequences of the parents of hybrid rice populations have been determined to support the search for an explanation of the molecular basis of heterosis in rice and tools to allow the prediction of hybrid performance. The sequencing of wild relatives from Africa is being used to better understand the utility of African wild rice varieties in crop improvement and the evolution of the A genome species of rice across Australia, Africa and South America.

Coffee

C. brassii grows to sea level and may be caffeine free, making this a useful species for the exploration of these traits by comparative genomics. Whole-genome comparative genomics has been established as a powerful tool for understanding key traits in the almond, cherry and peach (*Prunoideae*, Koepke *et al.*, 2013). The sequencing of domesticated coffee by an international consortium will deliver a reference sequence to allow the exploration of variation in wild relatives.

Macadamia

The four wild relatives of Macadamia in Southern Queensland and Northern New South Wales are the clade from which domesticated Macadamia is derived. Domesticated Macadamia is derived from *Macadamia integrifolia* and *Macadamia tetraphylla*, the two wild species that have edible fruit. The other two wild species are smaller trees that may be useful for breeding more compact trees. Chloroplast genome sequences have defined phylogenetic relationships within the genus. Further sequencing efforts are directed towards establishing a reference genome sequence for Macadamia as a tool for genetic improvement.

Eucalypts

Eucalypt genome sequencing has been carried out for several species. The re-sequencing of the large number of species in this group is a useful way to study the evolution of this group and to define where useful variation might be found to support genetic improvement of Eucalypts. The use of Eucalypts as a source of biomass to replace oil is being explored because of their ability to grow well in marginal environments where they are less likely to compete with food crops compared with many other species (Shepherd *et al.*, 2011). Association analysis of the relationships between bioenergy traits and whole-genome sequences is a strategy to rapidly advance this application (Henry, 2012a). The genus *Corymbia* comprises species that may be of special value as bioenergy crops. Current efforts aim to establish reference genome sequences for this new genus and facilitate the exploration of the variation in genomes in relation to bioenergy traits (Lupoi *et al.*, 2013).

Current status of technology for plant germplasm characterization

DNA sequencing platforms are continuing to improve at a rapid pace. The range of applications in plant germplasm characterization is expanding with sequencing becoming cheaper and easier. The ability to multiplex (combine multiple samples in a single lane or well for analysis) promises to make whole-genome shotgun sequencing a routine tool in the characterization of plant genetic resources. Direct comparisons of current DNA sequencing platforms suggest complementary capabilities and likely continuing improvement in suitability for application to the characterization of plant genetic resources. Current platforms and data analysis tools make whole-chloroplast genome analysis a simple starting point

Table 1. Wild plant samples subjected to sequencing

Species	Objectives
<i>Oryza</i> spp.	Characterizing the diversity of wild populations Establishing phylogenetic relationships among wild taxa Supporting the taxonomic revision of wild taxa Identifying germplasm useful for adapting rice crop to variable climates
<i>Coffea</i> spp.	Defining the genetics of unique domesticated traits through comparative genomics
<i>Macadamia</i> spp.	Determining relationships between wild species and domesticated varieties Defining the potential of wild species to contribute to genetic improvement
<i>Eucalyptus</i> and <i>Corymbia</i> spp.	Establishing phylogenetic relationships between wild species Identifying diversity in genes supporting the breeding of Eucalypts for new uses in energy and biomaterial production

providing a cost-effective universal barcode for plant germplasm. This is likely to extend to whole nuclear genomes as the technology improves over the next few years, especially for plants with smaller genomes and for which a reference genome of a close relative is available. The capture of diversity in wild material will be greatly accelerated by selection at the whole-genome level (Henry, 2012b). It is reasonable to expect sequencing costs to continue to drop eventually, making the sequencing of entire collections of plant genetic resources a useful tool for effective management of collections.

References

- Edwards D, Henry RJ and Edwards KJ (2012) Advances in DNA sequencing accelerating plant biotechnology. *Plant Biotechnology Journal* 10: 621–622.
- Fitzgerald TL, Shapter FM, McDonald S, Waters DLE, Chivers IH, Drenth A, Nevo E and Henry RJ (2011) Genome diversity in wild grasses under environmental stress. *Proceeding of the National Academy of Science USA* 108: 21139–21144.
- Gopala Krishnan S, Waters DLE, Katiyar SK, Sadananda AR, Satyadev V and Henry R (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal* 10: 623–634.
- Henry RJ (2012a) Advances in DNA sequencing enabling more rapid development of improved biomass and biofuel conversion technologies. *Biofuels* 3: 507–509.
- Henry RJ (2012b) Next generation sequencing for understanding and accelerating crop domestication. *Briefings in Functional Genomics* 11: 51–56.
- Henry RJ, Edwards M, Waters DLE, Krishnan SG, Bundock P, Sexton TR, Masouleh AK, Nock CJ and Pattermore J (2012) Application of large scale sequencing to marker discovery in plants. *Journal of Bioscience* 37: 829–841.
- Koepke T, Schaeffer S, Harper A, Dicenta F, Henry RJ, Moller BL, Sanchez-Perez R and Dhingra A (2013) Comparative genomics analysis in prunoideae to identify biologically relevant polymorphisms. *Plant Biotechnology Journal* 11, 883–893.
- Lupoi JS, Singh S, Simmons B and Henry R (2013) Assessment of lignocellulosic biomass using analytical spectroscopy: an evolution to high throughput techniques. *Bioenergy Research*. doi:10.1007/s12155-013-9352-1.
- Nock C, Waters DLE, Edwards MA, Bowen S, Rice N, Cordeiro GM and Henry RJ (2011) Chloroplast genome sequence from total DNA for plant identification. *Plant Biotechnology Journal* 9: 328–333.
- Persil-Cetinkol O, Smith-Moritz AM, Cheng G, Lao J, George A, Hong K, Henry R, Simmons B, Heazlewood JL and Holmes BM (2012) Structural and chemical characterization of hardwood from tree species with applications as bio-energy feedstocks. *PLoS ONE* 7: e52820.
- Shapter FM, Fitzgerald TL, Waters DLE, McDonald S, Chivers IH and Henry RJ (2012) Analysis of adaptive ribosomal gene diversity in wild plant populations from contrasting climatic environments. *Plant Signaling and Behavior* 7: 1–3.
- Shepherd M, Bartle J, Lee DJ, Brawner J, Bush D, Turnbull P, Macdonell P, Brown TR, Simmons B and Henry R (2011) Eucalypts as a biofuel feedstock. *Biofuels* 2: 639–657.
- Waters DLE, Nock CJ, Ishikawa R, Rice N and Henry RJ (2012) Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecology and Evolution* 2: 211–217.