



ASSESSING HERBARIUM MATERIAL WITH NOVEL MOLECULAR TECHNIQUES REVEALS A WEALTH OF NEW DATA FROM OLD TREASURE TROVES

by PAUL NEVILL

Curtin University, School of Molecular and Life Sciences, ARC Centre for Mine Site Restoration
Trace and Environmental DNA (TrEnD) Lab, Perth, Western Australia

Hamersley Range, Pilbara, Western Australia
PHOTO CREDIT: Stephen van Leeuwen

Using large-scale genome skimming to build a resilient resource for the future

Herbaria are valuable sources of extensive curated plant material that are important reference specimens for plant identification. These plant materials are now also accessible to genetic studies because of advances in high-throughput, next-generation sequencing (NGS) methods.

In our study, we conducted a large-scale applied assessment of one such NGS approach – genome skimming – and its ability to recover plastid and ribosomal genome sequences from a broad taxonomic sampling of herbarium material for the Western Australian flora. We sequenced 672 samples covering 21 families, 142 genera, and 530 named and proposed named species, and explored the impact of sample age, DNA concentration and quality, read depth and fragment length on

plastid assembly error. We demonstrate that herbaria are a valuable source of plant material for building a comprehensive DNA sequence database which serves various applications from modernizing plant surveys to improving the resolution of plant phylogenies.

Genome skimming¹ was effective at producing genomic information at large scale. Substantial sequence information on the chloroplast genome was obtained from 96.1% of samples, and complete or near-complete sequences of the nuclear ribosomal RNA gene repeat were obtained from 93.3% of samples. We extracted sequences for plastid markers *rbcL* and *matK* – the core DNA barcode regions – from 96.4% and 93.3% of samples, respectively. Read quality and DNA fragment length had significant effects on sequencing

outcomes and error correction of reads proved essential. Assembly problems were specific to certain taxa with low GC and high repeat content (e.g. *Goodenia*, *Scaevola*, *Cyperus*, *Bulbostylis*, *Fimbristylis*), suggesting the influence of biological rather than technical factors. The structure of related genomes was needed to guide the assembly of repeats that exceeded the read length. DNA-based matching proved highly effective and showed that the efficacy for species identification declined in the following order: total chloroplast DNA >> ribosomal DNA > matK >> rbcL.



Eucalyptus kingsmillii, Pilbara, Western Australia

PHOTO CREDIT: Stephen van Leeuwen



Gastrolobium grandiflorum, Pilbara, Western Australia

PHOTO CREDIT: Stephen van Leeuwen

Our success is important as it demonstrates that herbaria can be used as a source of plant material for building a comprehensive DNA sequence database. These data form the basis of development of a molecular identification system for the Western Australian flora. This will enable identification of specimens throughout the year (e.g., non-flowering times) and for hard-to-identify species (e.g., those with constrained or reduced morphological characters) or for speci-

mens where only fragments of non-diagnostic material are available. The availability of this technology will modernize plant surveys by reducing constraints on survey effort through moderating sampling timing restrictions and seasonal effects, as well as enabling rapid verifiable identification. It will also have practical applications in a wide range of ecological contexts using eDNA metabarcoding, such as gut and scat analysis of animals to determine dietary preferences of threatened species and livestock, and checking the integrity of seed collections for seed banking and use in land restoration/revegetation programs. Other potential uses of extensive plastid sequence data, beyond species identification, include improving the resolution of plant phylogenies and studies on the evolution of plastid genome function, including understanding adaptive changes.

For full details, please refer to the publication:

Nevill PG et al. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. *Plant Methods* 16, 1 (2020). <https://doi.org/10.1186/s13007-019-0534-5>



Ptilotus rotundifolius, Pilbara, Western Australia

PHOTO CREDIT: Stephen van Leeuwen

References

1. Straub S, Parks M, Weitemier K, Fishbein M, Cronn R, Liston A (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99(2), 349-364. <https://dx.doi.org/10.3732/ajb.1100335>

Online:

<https://ibol.org/barcodebulletin/research/assessing-herbarium-material-with-novel-molecular-techniques-reveals-a-wealth-of-new-data-from-old-treasure-troves/>