The Genomics for Australian Plants (GAP) framework initiative – developing genomic resources for understanding the evolution and conservation of the Australian flora.

Lalita Simpson¹, David J. Cantrill²,³, Margaret Byrne⁴, Theodore R. Allnutt⁵, Graham J. King⁶, Mabel Lum⁷, Ziad Al Bkhetan⁸, Rose L. Andrew⁹, William J. Baker¹, İ. Matthew D. Barrett¹, Jacqueline Batley¹, Oliver Berry¹, Rachel M. Binks⁰, Jason Bragg¹⁰, Linda Broadhurst¹, Gillian Brown⁰, Jeremy J. Bruhl¹¹, Richard J. Edwards¹², Scott Ferguson⁰, Félix Forest¹, Johan Gustafsson⁰, Timothy A. Hammer¹³,⁴, Gareth D. Holmes⁵, Christopher J. Jackson⁸, Elizabeth A. James⁸, Ashley Jones⁸, Paul J. Kersey¹, Ilia J. Leitch¹, Olivier Maurin¹, Todd G. B. McLay¹, Daniel J. Murphy¹,³,⁵, Katharina Nargar²,³,⁴, Lars Nauheimer¹, Hervé Sauquet¹⁵, Alexander N. Schmidt-Lebuhn⁴, Kelly A. Shepherd¹⁶, Anna E. Syme⁶, Michelle Waycott¹³,⁴, Trevor C. Wilson⁷, Darren M. Crayn¹,⁵

Correspondence to: Darren M. Crayn, Australian Tropical Herbarium, James Cook University Nguma Bada Campus, Smithfield, Qld 4878, Australia. Email: darren.crayn@jcu.edu.au

¹Australian Tropical Herbarium, James Cook University Nguma Bada Campus, Smithfield, Qld 4878, Australia
²Royal Botanic Gardens Victoria, Melbourne, VIC 3004, Australia
³School of BioSciences, The University of Melbourne, Parkville, VIC 3010, Australia
⁴Biodiversity and Conservation Science, Department of Biodiversity, Conservation and Attractions, Locked Bag 104, Bentley Delivery Centre, Bentley, WA 6983, Australia
⁵Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia
⁶Bioplatforms Australia Ltd., Sydney, NSW 2109, Australia
⁷Australian Biocommons, The University of Melbourne, Parkville, VIC 3010, Australia
⁸Botany & N.C.W. Beadle Herbarium, University of New England, Armidale, NSW 2351, Australia
⁹Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom
¹⁰Department of Biology, Aarhus University, Ny Munkegade 116, 8000 Aarhus C, Denmark
¹¹School of Biological Sciences and Institute of Agriculture, University of Western Australia, Crawley, WA 6009, Australia
¹²CSIRO Environomics Future Science Platform, Indian Ocean Marine Research Centre, University of Western Australia, Crawley, WA 6009, Australia
¹³Research Centre for Ecosystem Resilience, Botanic Gardens of Sydney, Sydney, NSW 2000, Australia
¹⁴Centre for Australian National Biodiversity Research (a joint venture of Parks Australia and CSIRO), GPO Box 1700, Canberra, ACT 2601, Australia
¹⁵Queensland Herbarium and Biodiversity Science, Department of Environment and Science, Brisbane Botanic Gardens, Toowong, Qld 4066, Australia
¹⁶Minderoo OceanOmics Centre at UWA, Oceans Institute, University of Western Australia, Perth, WA 6009, Australia
¹⁷Evolution and Ecology Research Centre, School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, NSW 2052, Australia
¹⁸Research School of Biology, Australian National University, Canberra, ACT 0200, Australia
¹⁹University of Adelaide, Adelaide, SA 5005, Australia
²⁰Botanic Gardens and State Herbarium, North Terrace, Adelaide, SA 5000, Australia
²¹National Biodiversity DNA Library, CSIRO, Parkville, VIC 3010, Australia
²²Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia
²³National Research Collections Australia, Commonwealth Industrial and Scientific Research Organisation (CSIRO), GPO Box 1700, Canberra, ACT 2601, Australia
²⁴National Herbarium of NSW, Botanic Gardens of Sydney, Mount Annan, New South Wales, Australia
Abstract

The generation and analysis of genome-scale data—genomics—is driving a rapid increase in plant biodiversity knowledge. However, the speed and complexity of technological advance in genomics presents challenges for its widescale use in evolutionary and conservation biology. Here, we introduce and describe a national-scale collaboration conceived to build genomic resources and capability for understanding the Australian flora: the Genomics for Australian Plants (GAP) Framework Initiative. We outline (a) the history of the project including the collaborative framework, partners, and funding; (b) GAP principles such as rigour in design, sample verification and documentation, data management, and data accessibility; and (c) the structure of the consortium and its four activity streams (reference genomes, phylogenomics, conservation genomics, and training), with the rationale and aims for each of them. We show, through discussion of its successes and challenges, the value of this multi-institutional consortium approach and the enablers, such as well-curated collections and national collaborative research infrastructure, all of which have led to a substantial increase in capacity and delivery of biodiversity knowledge outcomes.

Keywords: phylogenomics, population genetics, reference genomes, Angiosperms353, systematics, taxonomy

Introduction

Australia is a large and geologically relatively stable landmass that is home to 22,600 accepted vascular plant species (ca. 6.3% of the world’s flora; Govaerts et al. 2023) in 2144 genera (Council of Heads of Australasian Herbaria 2024). This flora is well known for its diversity (>225 families) and endemism (>90% of species, Chapman 2009) and occurs across a wide range of biomes including the mesic south-east and south-west coasts to monsoon tropics across the north, and the large arid zone that dominates the inland part of the continent (Byrne et al. 2008, 2011, Bowman et al. 2010).

Documenting the diversity and relationships of the Australian flora has been a focus from the early days of plant collecting, while more recently molecular systematics has been employed in an integrated approach to understanding phylogenetic relationships, particularly at genus and family levels. Indeed, molecular systematics has highlighted the role of isolation and species radiations as important features in the evolutionary history of much of the arid and mesic flora of Australia (Byrne et al. 2008, 2011, Crisp and Cook 2013), with immigration being a stronger feature in the monsoonal flora through its proximity to south-east Asia (Crayn et al. 2015). At lower taxonomic levels, phylogeography has revealed a strong signal for persistence of species through Pleistocene climatic changes, providing insights into current species distributions (Byrne et al. 2008, 2011, Byrne and Murphy 2020), while population genetics has strengthened our understanding of the significance of diversity for conservation (Broadhurst et al. 2017).

Much of the taxonomic and systematic research to document and describe the Australian flora and understand its relationships has been undertaken at herbaria and botanic gardens, as these institutions are centres for botanical collections and knowledge. Since the first local botanic institution (and Australia’s first scientific institution) was established in 1816 (Botanic Gardens of Sydney), Australia has developed 31 recognised national, state and university vascular plant herbaria (Atlas of Living Australia “collectory” collections.ala.org.au, Thiers 2023), and 139 botanic gardens. Australia’s major herbaria were established by governments (State, Territory and Commonwealth), principally to research the flora and to provide botanical information and advice. Many universities have developed herbaria to support research and education.
There has been a long history of collaboration among herbaria nationwide, which became formalised in 1972 with the establishment of the Council of Heads of Australasian Herbaria (CHAH). Enabled by CHAH, Australian herbaria have been at the forefront globally of plant biodiversity science and data delivery, exemplified by several major knowledge infrastructure projects and initiatives including a national specimen databasing initiative, Australia’s Virtual Herbarium (AVH – www.avh.ala.org.au), which later spawned the Atlas of Living Australia (ALA – www.ala.org.au), the Australian Plant Names Index (APNI – https://biodiversity.org.au/nsl/services/) and the Australian Plant Census (APC – https://biodiversity.org.au/nsl/services/), which is a national consensus checklist of accepted plant names.

In addition to leading in collections data and knowledge delivery, researchers at Australian herbaria and botanic gardens have been early adopters of molecular tools in systematics and evolution research. Nevertheless, the speed and complexity of technological advance in molecular genomics and the concomitant challenges in data generation, analysis, management, and use, have outpaced the capacity of many institutions and individual researchers to utilise these tools optimally. The Genomics for Australian Plants (GAP) Framework Initiative is a consortium established to strategically address this gap and foster implementation of genomics technology to build plant knowledge across the research and wider community. In this paper we outline the principles and approach taken in establishing this consortium, describe the aims and architecture of the research projects undertaken, and detail progress to date against these aims. Several major milestones have been achieved so far: 34 reference genomes sequenced, a nuclear sequence data resource for 90% of Australian native angiosperm genera with an accompanying phylogenomic tree (Australian Angiosperm Tree of Life, AAToL), and the resolution of 10 species complexes of conservation concern. These milestones clearly demonstrate that a consortium approach enables the plant biodiversity science community to achieve the scale required for effective national capacity building, utilisation of novel genomics technologies, and delivery of large-scale datasets and knowledge for understanding the evolution and conservation of the Australian flora.

Development of the Genomics for Australian Plants (GAP) Framework Initiative

Recognising the opportunity to enhance genomics research capability and infrastructure for the plant biodiversity sciences, Bioplatforms Australia (https://bioplatforms.com/) initiated discussions in 2016 with the Australian herbarium and botanic gardens communities, and internationally with Royal Botanic Gardens Kew, to explore the development of a consortium to undertake a collaborative, capacity-building genomics project.

Drawing on the findings from the Oz Mammals Genomics Framework Data Initiative (Eldridge et al. 2020), a draft proposal for the development of the consortium was circulated to Australian herbaria, interested individuals working in botanic gardens and universities, and subject-matter experts in other organisations. Participants at an initial workshop in Melbourne in August 2017 agreed to further develop, through subsequent meetings, the consortium model, which led to the establishment of a Steering Committee and two Working Groups to advise the Steering Committee on strategy and methodological approaches: the Wet Lab Working Group and the Computational Working Group. The GAP Framework Initiative was formally launched at the Australasian Systematic Botany Society meeting in Brisbane in December 2018.

The development of the GAP Framework Initiative was undertaken in the context of, and enabled by, the resources, infrastructure, botanical knowledge, and genomic expertise of the Australian national herbarium network. The GAP Framework Initiative is also aligned with commitments outlined in national strategic plans such as Australia’s Strategy for Nature (Interjurisdictional Biodiversity Working Group, 2019) and the Decadal Plan for Taxonomy and Biosystematics (Taxonomy Decadal Plan Working Group, 2018). The consortium model uses existing resources to enable and facilitate the development and adoption of innovative technologies to generate data, as well as maximise the uptake and use of outputs (Fig. 1). Details of inputs, activities, outputs, and outcomes are outlined below.
**Inputs (data, people, and institutional stakeholders) are the infrastructure that enables GAP data-generating activities that are undertaken in three streams: reference genomes, phylogenomics, and conservation genomics. Training activities are embedded within, and link across, all three data streams. Together, these activities produce outputs and outcomes that enhance and enrich the human, knowledge, and institutional resources that enabled the project.**

**Inputs**
The infrastructure that enables the GAP Framework Initiative comprises people (a well-trained research workforce), knowledge (findable, accessible, interoperable and reusable data — FAIR; Wilkinson et al. 2016), and physical elements (world-class biological collections, laboratories, instruments, and computational resources) (Fig. 1). Core funding was provided by Bioplatforms Australia (funded by the Australian Government’s National Collaborative Research Infrastructure Strategy fund - NCRIS), philanthropy (Ian Potter Foundation, Royal Botanic Gardens (Victoria) Foundation), and institutional partners (herbaria and botanical gardens). Further investment was provided through the contributions of the more than 210 systematists, taxonomists, herbarium and laboratory curators and technicians, project coordinators, bioinformaticians, ecologists, and conservation scientists that comprise the GAP consortium. Collectively, these individuals represent 28 organisations, including government agencies, herbaria, botanical gardens, universities and genomics laboratories from all states and territories. Plant samples were sourced from
curated botanical collections maintained in herbaria and botanical gardens, supplemented with wild collected material where required. Sample metadata curation and management were verified with existing biodiversity data infrastructure such as the AVH/ALA, and National Species List (APNI/APC). Generation and management of genomic data is supported by Bioplatforms Australia’s genomic facilities including the ACRF Biomolecular Resource Facility (BRF, https://jcsmr.anu.edu.au/research/facilities/brf), the Australian Genomics Research Facility (AGRF, https://www.agrif.org.au/), Genomics Western Australia (GWA, https://www.genomicswa.com.au/), the Ramaciotti Centre for Genomics (https://www.ramaciotti.unsw.edu.au/), and the Australian BioCommons (https://www.biocommons.org.au/). Computational resources for data analysis are provided through the Australian BioCommons and partners such as the National Computational Infrastructure (NCI, https://nci.org.au/). GAP partners and core funding providers are listed on the GAP website (https://www.genomicsforaustralianplants.com/consortium/).

GAP Aims and Principles

Aims

The GAP Framework Initiative aimed to bring together researchers and data specialists in plant science and conservation organisations to increase and share biodiversity knowledge, accelerate species discovery, enhance biodiversity collections, and support strategic growth of taxonomic and biosystematic capacity and capabilities. The aims have been achieved by:

- Generating plant genomic data across the plant tree of life to enable better conservation, utilisation and understanding of Australia’s unique plant diversity;
- Building genomic capacity across Australian botanic gardens and herbaria to create networks that can collaborate on the collection, management, dissemination and application of genomic data for Australian plants;
- Providing tools that enable the use of genomic data for identification and classification of biodiversity at a range of scales, and to use these tools to inform conservation management and enable better decision making; and
- Upskilling the plant science community in the use of new sequencing technologies through the development of wet lab skills and bioinformatics training modules.

Principles

The GAP Framework Initiative has operated under a set of research and data standards as follows:

1. The taxonomic identity of all plant material used in the GAP Framework Initiative is determined by a taxonomic specialist.
2. Taxonomy conforms to the APC or recent peer reviewed taxonomic publications not yet considered by the APC.
3. Plant material was collected in Australia (including offshore territories).
4. Plant material was collected legally and is compliant with the Nagoya Protocol1.
5. Plant material was vouchered by specimens lodged in a recognised herbarium (Index Herbariorum, Thiers 2023).
6. Samples comply with rigorous metadata standards that include herbarium voucher accession and genomic sample processing details.

---

1 The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity is an international agreement that aims to ensure that the benefits arising from the utilization of genetic resources are shared in a fair and equitable way. It entered into force on 12 October 2014. While not yet ratified in Australia, in practice, GAP participants demonstrate compliance with the spirit of the Nagoya Protocol by identifying the owner of the genetic material submitted for analysis. For samples obtained from public land, the owner is evidenced by collecting or scientific research permits from the relevant authority. In the case of samples collected on Indigenous or other land including land in private ownership, evidence may comprise a written agreement that demonstrates prior, informed consent was given by the landowner for genetic research to be conducted on the sample(s).
7. All data are made available to consortium partners immediately through the Bioplatforms Australia Data Portal (https://data.bioplatforms.com/) and following data publication are made freely accessible through public repositories (e.g. the European National Archive - ENA) in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson et al. 2016). Software, workflows and training materials are also made freely available.

8. Individual plants (or for outcrossing annuals, an individual of the same population or the closest heritable unit to the sampled ortet (ramet)) sampled for reference genome projects are conserved in living collections to enable future research projects to access the same individual or lineage.

9. For phylogenomics projects, type species of genera are sampled where possible.

GAP Activities

The GAP consortium has undertaken data generation, workflow development and capacity building across four activity streams:

1. Reference genomes;
2. Phylogenomics;
3. Conservation genomics;
4. Training.

Activity 1: Reference Genomes

While recent years have seen a marked acceleration in the availability of plant nuclear genomic resources, the great majority of these are for crop species, their wild relatives, and forest trees. Thus, resources are very unevenly distributed across the plant tree of life, which hampers their use to understand patterns of genomic diversity and processes that drive genome evolution. The GAP Reference Genomes stream aimed to address this resource bias for Australian lineages of flowering plants (angiosperms). Initially, a Working Group was formed to consider: 1) which of the increasing array of short- and long-read genomic sequencing technologies are most appropriate for the task; and 2) how to prioritise taxa for sequencing.

To identify orders and families in the Australian angiosperm flora that lacked genomic resources, the GAP Steering Committee led the compilation of a list of all plant nuclear genomic resources available from public repositories (NCBI genomes, NCBI Sequence Read Archive, PlaBiPD), including completed genomes, draft genomes / assemblies, and other sequence data (e.g. RNAseq, genome skimming). This list is available online (see links under GAP Data and Resources Availability section below). Using this list for priority setting, community participation in the GAP Reference Genomes stream was invited through requests for partnership in four phases. This approach enabled the consortium to adapt the program in response to new technologies as they became available and to progressively tackle larger genomes. Plants with diploid genomes between 1–2 Gigabases (Gb; 1C-value) were targeted as these were considered tractable for sequencing to sufficient depths for assembly. Additionally, the lower cost of sequencing for genomes in this size range compared with larger genomes meant that the GAP Framework Initiative could sequence more representatives to fill taxonomic gaps across the tree of life.

Extracting high molecular weight DNA suitable for reference genome assembly was a significant challenge for nearly all taxa, and exceptionally difficult for some (Jones et al. 2021). To address this difficulty, members of the consortium developed dedicated protocols suitable for extracting high molecular weight DNA for long-read sequencing of diverse non-model plants. This included virtual engagement with the wider scientific community for interactive protocol development, discussion, and open access sharing through Protocols.io (Teytelman et al. 2016). Nevertheless, several projects could not be completed because obtaining DNA of sufficient quality and quantity proved impossible. All of the species for which
nuclear genomes were sequenced through the GAP Framework Initiative are listed in Table 1, including those that are not yet completed at time of publication.

Table 1. List of the GAP nuclear reference genome projects undertaken, including the taxon sampled, genome size, sequencing method used, the current status of the project, outputs based on the sequence data, and location of living voucher plant. Sequencing methods are as follows: 10x = 10x Genomics (Pleasanton, CA, USA), ONT = Oxford Nanopore Technology, a long read single molecule sequencing approach (Oxford Nanopore, Oxford, UK), Hi-C = a genomic technique that captures chromatin conformation (Lieberman-Aiden et al. 2009), HiFi = high fidelity long read sequencing approach (PacBio, Menlo Park, CA, USA), Illumina = short read sequencing approach (Illumina, San Diego, CA, USA). All species are diploid except *Denhamia bilocularis* (F.Muell.) M.P.Simmons (polyploid, ploidy level unknown), and *Flindersia xanthoxyla* (A.Cunn. ex Hook.) Domin (tetraploid).  

- A Indicates orders and families (APG IV; The Angiosperm Phylogeny Group 2016) for which no published genomic resources existed at the start of this project.  
- B Indicates species for which genome size was determined experimentally using flow cytometry.  
- C Indicates species for which genome size was estimated based on relatives in the order (Pellicer and Leitch 2019).  
- D Indicates species for which genome size was estimated based on relatives in the family (Pellicer and Leitch 2019).  
- E Indicates species for which genome size was estimated based on relatives in the genus (Pellicer and Leitch 2019). All other genome size estimates were sourced from the Plant DNA C-values Database Release 7.1 (Pellicer and Leitch 2019). Locations of living vouchers are as follows: ABG = Adelaide Botanic Gardens, Adelaide SA; ANBG = Australian National Botanic Gardens, Canberra ACT; KPBG = Kings Park Botanic Garden, Perth WA; RBGV = Royal Botanic Gardens Victoria, Melbourne VIC; RGV-C = Royal Botanic Gardens Victoria, Cranbourne VIC; CBotG = Cairns Botanic Gardens, Cairns QLD; NENP = New England National Park, NSW; UMelb = Systems Garden, University of Melbourne, Melbourne VIC; UNE = University of New England, Armidale NSW; WP = Wellington Park, TAS.
<table>
<thead>
<tr>
<th>Order</th>
<th>Family</th>
<th>Species</th>
<th>Genome size estimate (Gb, 1C-value)</th>
<th>Sequencing Method</th>
<th>Project Status</th>
<th>Outputs</th>
<th>Living voucher location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arecales</td>
<td>Dasypogonaceae&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Kinjia australis R.Br.</td>
<td>0.5&lt;sup&gt;0&lt;/sup&gt;</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>KPBG</td>
</tr>
<tr>
<td>Asparagales</td>
<td>Asphodelaceae</td>
<td>Xanthorrhoea johnsonii A.T.Lee</td>
<td>1.5&lt;sup&gt;B&lt;/sup&gt; (T. McLay, unpubl.)</td>
<td>HiFi, ONT, Hi-C, Illumina</td>
<td>Assembled</td>
<td>Ferguson et al. 2022</td>
<td>ANBG</td>
</tr>
<tr>
<td>Asparagales</td>
<td>Doryanthaceae&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Doryanthus excelsa Correa</td>
<td>2.94 (Smarda et al. 2014)</td>
<td>HiFi, ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Asparagales</td>
<td>Orchidaceae</td>
<td>Thelymitra variegata (Lindl.) F.Muell.</td>
<td>-</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>KPBG</td>
</tr>
<tr>
<td>Asterales</td>
<td>Asteraceae</td>
<td>Leucocrysum albicans (A.Cunn.) Paul G. Wilson</td>
<td>0.9&lt;sup&gt;B&lt;/sup&gt; (A. Schmidt-Lebuhn, unpubl.)</td>
<td>HiFi</td>
<td>In assembly</td>
<td>n/a</td>
<td>ANBG</td>
</tr>
<tr>
<td>Asterales</td>
<td>Campanulaceae</td>
<td>Wahlenbergia ceracea Lothian</td>
<td>0.3&lt;sup&gt;E&lt;/sup&gt;</td>
<td>HiFi, ONT, Illumina</td>
<td>Assembled</td>
<td>n/a</td>
<td>Unknown</td>
</tr>
<tr>
<td>Asterales</td>
<td>Goodeniaceae</td>
<td>Dampiera purpurea R.Br.</td>
<td>0.5&lt;sup&gt;E&lt;/sup&gt;</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Austrobaileyales&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Austrobaileyaceae&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Austrobaileyya scandens C.T.White</td>
<td>9.3 (Leitch and Hanson 2002)</td>
<td>HiFi, ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>CBotG</td>
</tr>
<tr>
<td>Berberidopsidales&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Berberidopsidaceae&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Berberidopsis beckleri (F.Muell.) Veldkamp</td>
<td>0.3&lt;sup&gt;E&lt;/sup&gt;</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>NENP</td>
</tr>
<tr>
<td>Brassicales</td>
<td>Cleomaceae</td>
<td>Areocleome oxalidea (F.Muell.) R.L.Barrett &amp; Roalson</td>
<td>-</td>
<td>Project ended</td>
<td>Project ended – failure to culture plant material</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Celastrales</td>
<td>Celastraceae</td>
<td>Denhamia bilocularis (F.Muell.) M.P.Simmons</td>
<td>0.8&lt;sup&gt;D&lt;/sup&gt;</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Dilleniales&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Dilleniaceae&lt;sup&gt;A&lt;/sup&gt;</td>
<td>Hibbertia scandens (Willd.) Gilg</td>
<td>2.64 (Garcia et al. 2010)</td>
<td>ONT, Illumina</td>
<td>In assembly</td>
<td>n/a</td>
<td>ABG</td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td>Acacia pycnantha Benth.</td>
<td>0.6&lt;sup&gt;B&lt;/sup&gt; (McLay et al. 2022)</td>
<td>ONT, Illumina</td>
<td>Completed</td>
<td>Syme et al. 2021 McLay et al. 2022</td>
<td>ANBG</td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td>Gastrolobium racemosum (Turcz.) Crisp</td>
<td>0.85&lt;sup&gt;A&lt;/sup&gt; (Schmidt-Lebuhn and Cantrill, 2023)</td>
<td>ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td>Pultenaea daphnoides J.C.Wendl.</td>
<td>0.7 (Morgan and Westoby 2005)</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Fagales</td>
<td>Casuarinaceae</td>
<td>Gymnostoma australianum L.A.S.Johnson</td>
<td>0.7&lt;sup&gt;D&lt;/sup&gt;</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Division</td>
<td>Family</td>
<td>Genus and Species</td>
<td>C</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>Location</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>-------------------------------------------</td>
<td>----</td>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>Fagales</td>
<td>Nothofagaceae</td>
<td>Nothofagus cunninghamii (Hook.) Oerst.</td>
<td>1</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Geraniales</td>
<td>Geraniaceae</td>
<td>Pelargonium australe Willd.</td>
<td>0.32</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>ABG</td>
</tr>
<tr>
<td>Lamiales</td>
<td>Scrophulariaceae</td>
<td>Eremophila drummondii F.Muell.</td>
<td>0.8 (A. Schmidt-Lebuhn and R. Fowler unpubl.)</td>
<td>HiFi, Illumina</td>
<td>Assembled</td>
<td>n/a</td>
<td>UMelb</td>
</tr>
<tr>
<td>Laurales</td>
<td>Monimiaceae</td>
<td>Hedycarya angustifolia A.Cunn.</td>
<td>0.7 (Schmidt-Lebuhn and Cantrill, 2023)</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Magnoliales</td>
<td>Eupomatiaceae</td>
<td>Eupomatia laurina R.Br.</td>
<td>1.2</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Malpighiales</td>
<td>Phyllanthaceae</td>
<td>Phyllanthus gunnii Hook.f.</td>
<td>1.0 (Schmidt-Lebuhn and Cantrill, 2023)</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Myrtales</td>
<td>Myrtaceae</td>
<td>Archirhodomyrtus beckleri (F.Muell.) A.J.Scott</td>
<td>0.8</td>
<td>HiFi, ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Oxalidales</td>
<td>Cunoniaceae</td>
<td>Callicoma serratifolia Andrews</td>
<td>-</td>
<td>ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Oxalidales</td>
<td>Cunoniaceae</td>
<td>Eucryphia lucida (Labill.) Baill.</td>
<td>-</td>
<td>ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Oxalidales</td>
<td>Elaeocarpaceae</td>
<td>Elaeocarpus reticulatus Sm.</td>
<td>0.3</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Paracryphiales</td>
<td>Paracryphiaceae</td>
<td>Quintinia fawkneri F.Muell.</td>
<td>1.2 (Schmidt-Lebuhn and Cantrill, 2023)</td>
<td>HiFi, ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGC</td>
</tr>
<tr>
<td>Poales</td>
<td>Cyperaceae</td>
<td>Lepidosperma gladiatum Labill.</td>
<td>0.6</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Proteales</td>
<td>Proteaceae</td>
<td>Bellendena montana R.Br.</td>
<td>4.86</td>
<td>HiFi</td>
<td>In assembly</td>
<td>n/a</td>
<td>WP</td>
</tr>
<tr>
<td>Proteales</td>
<td>Proteaceae</td>
<td>Isopogon anethifolius (Salisb.) Knight</td>
<td>0.6</td>
<td>ONT</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Proteales</td>
<td>Proteaceae</td>
<td>Telopea speciosissima (Sm.) R.Br.</td>
<td>0.9 (Chen et al. 2022)</td>
<td>ONT, 10x, Hi-C</td>
<td>Completed</td>
<td>Chen et al. 2022</td>
<td>n/a (plant died)</td>
</tr>
<tr>
<td>Rosales</td>
<td>Urticaceae</td>
<td>Dendrocnide excelsa (Wedd.) Chew</td>
<td>1.6</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Sapindales</td>
<td>Rutaceae</td>
<td>Flindersia xanthoxyla (A.Cunn. ex Hook.) Domin</td>
<td>0.4 (Schmidt-Lebuhn and Cantrill, 2023)</td>
<td>HiFi</td>
<td>Assembled</td>
<td>n/a</td>
<td>RBGM</td>
</tr>
<tr>
<td>Sapindales</td>
<td>Rutaceae</td>
<td>Phebalium stellatum I.Telford &amp; J.J.Bruhl</td>
<td>1.5</td>
<td>HiFi, ONT, Illumina</td>
<td>Assembled</td>
<td>Ferguson et al. 2022</td>
<td>UNE</td>
</tr>
</tbody>
</table>
Results

To date, the GAP Reference Genomes stream has delivered: (1) new protocols for high-molecular weight DNA extraction, clean-up and size selection for long-read sequencing (Jones et al. 2021; Jones and Schwessinger 2023); (2) an approach to base-caller selection to improve actual accuracy of nanopore sequencing (Ferguson et al. 2022); (3) case studies in the use of long-read assemblies to reveal structural diversity in organellar genomes (Syme et al. 2021); (4) genome size estimates for 10 species (Chen et al. 2022; Mc Lay et al. 2022; Schmidt-Lebuhn and Cantrill 2023; T. McLay unpubl., A. Schmidt-Lebuhn and R. Fowler unpubl.); (5) an online concatenated database of published genomic resources compiled from public repositories; and (6) published genomic resources for Australia’s national floral emblem, the Golden Wattle, *Acacia pycnantha* Benth. (Syme et al. 2021; McLay et al. 2022), and the state floral emblem of New South Wales, the Waratah, *Telopea speciosissima* (Sm.) R.Br. (Chen et al. 2022).

The status of all 34 reference genome projects is provided in Table 1. Together, these 34 projects provide genomic resources for 21 orders and 27 families, and for four of these orders (Austrobaileyales Takht. ex Reveal, Berberidopsidales Doweld, Dilleniataes DC. ex Bercht. & J.Presl, Paracryphiales Takht. ex Reveal) and nine of these families, no published genomic resources exist (i.e. Austrobaileyaceae Croizat, Berberidopsidaceae Takht., Cunoniaceae R.Br., Dasygononaceae Dumort., Doryanthaceae R.Dahlgren & Clifford, Dilleniataes Salisb., Eupomatiaceae Orb., Monimiaceae Juss., Paracryphiaceae Airy Shaw). Most of the genome projects that are nearing completion will be published in a single overarching and descriptive GAP genome paper.

Activity 2: Phylogenomics – The Australian Angiosperm Tree of Life (AAToL)

An understanding of phylogenetic relationships underpins all evolutionary biology, including systematics. To date, plant molecular phylogenetic research in Australia has been largely directed by individual researcher or research institution needs and interests, rather than as a nationwide coordinated effort, with different approaches to marker choice and sampling across groups limiting the interoperability and reusability of data. The GAP Phylogenomics activity stream aimed to transform phylogenetic research on the Australian flora by producing a sequence data resource of common markers and by making available a phylogenetic tree containing at least 95% of Australian native angiosperm genera (i.e. 2037/2144 total genera, APC accessed 28 June 2024) – the Australian Angiosperm Tree of Life (AAToL). Progress has been achieved by supporting and coordinating collaborators in achieving their research goals where they contribute to building the AAToL.

Two working groups were formed to design and coordinate the GAP phylogenomics component. The Phylogenomics Working Group conceived the project aims and design, engaged the community, and coordinated the sampling strategy. The Phylogenomics Bioinformatics Working Group provided guidance and recommendations for approaches to data analysis, development of bioinformatic tools for analysis of target sequence capture datasets, generation of the AAToL tree, and offered bioinformatics support.

The Phylogenomics Working Group advised the adoption of target capture (also known as hybrid enrichment, target enrichment, and exon capture) as the sequencing approach, which offers reliable retrieval of hundreds or thousands of loci from across the nuclear genome (Cronn et al. 2015). A decision was then made to engage with the Plant and Fungal Trees of Life (PAFTOL; Baker et al. 2022) project led by the Royal Botanic Gardens Kew since 2015, which aims to build the tree of life for all angiosperm genera. Given their aligned approach, GAP and PAFTOL agreed to work in close partnership toward resolving the AAToL, with the GAP team reciprocating by making a subset of data available to the PAFTOL project. This has been achieved by coordinating sampling and methods for genomic data generation, as well as collaboration on scientific research outputs.
The aim was to construct the AAToL to genus level in two stages with community participation invited through Requests for Partnership:

**Stage 1** - In collaboration with PAFTOL, resolve the AAToL to genus level using a single exemplar species to ensure representation for >95% of accepted Australian genera.

**Stage 2** - Generate phylogenomic datasets densely sampled within selected genera, families, or orders to address questions of monophyly, evolution, and biogeography.

**Project Stages**

**Pilot study**

A pilot was undertaken to test the suitability of the Angiosperms353 (Johnson et al. 2019, Baker et al. 2021) universal target capture bait set for resolving the AAToL and to refine sampling and data workflows at a national scale. Through a call for partnership, participants were invited to share existing DNA libraries for analysis. Six responses were received offering libraries representing 166 species from 134 genera and 67 families. Sequencing was carried out following the GAP phylogenomics protocol (see www.genomicsforaustralianplants.com/wp-content/uploads/2022/10/GAP-Phylogenomics-protocols_Oct2022.pdf)

**The Australian Angiosperm Tree of Life (AAToL)**

**Stage 1**

Initially, the Phylogenomics Working Group scoped the project by determining those genera in the Australian native angiosperm flora (based on the APC; Council of Heads of Australasian Herbaria 2023) for which suitable phylogenomics data from a sample collected in Australia was available, either from the GAP pilot study, the PAFTOL project (which had already sequenced more than 600 Australian genera) or other projects such as One Thousand Plant Transcriptomes Initiative (1KP; Leebens-Mack et al. 2019). Six teams responded to the call for partnership, each nominating to sample a selection of the remaining unsampled genera based on the research interests of their team members, institutional capability, and availability of suitable samples. These six teams were led by members of the six largest Australian herbaria - AD, BRI, CANB, MEL, NSW and PERTH - and included members of five other herbaria - CNS, DNA, HO, MELU, UNE (abbreviations per Index Herbariorum, Thiers 2023). Collectively, all states and territories were represented. Together, the six teams brought the total number of samples included in GAP stage 1 to 1861 (87% of genera).

**Stage 2**

Sixteen requests for partnership for the second stage of AATOL were accepted and projects initiated. These projects proposed to undertake dense taxonomic sampling within monophyletic groups (orders to genera) (Table 2). Based on results from Stage 1 (data not shown) and from other studies (e.g. Baker et al. 2021), Angiosperms353 baits were expected to underperform for certain groups and consequently, GAP supported the use of additional bait sets. Thus, the OZbaits (Waycott et al. 2021) set was used for Alismatales R.Br. ex Bercht. & J.Presl, Chamaelauriaceae DC. (Myrtaceae Juss.), Drosera L. (Droseraceae), Hibbertia Andrews (Dilleniaceae), Lasiopetaleae Gay (Malvaceae Juss.), Santalaceae R.Br., and Lazarum A.Hay (Araceae Juss.), and the SaliBaits set (A. Žerdoner Čalasan, unpublished) was used for Tecticornia Hook.f./Salicornia L. (Chenopodiaceae Vent.) (Table 2). In addition to the bespoke bait set data, Angiosperms353 data was generated for all samples in these projects to maximise combinability and reuse of datasets and to enable head-to-head comparisons of bait set performance to provide evidence for bait choice and future bait design development.
Table 2. List of the GAP Phylogenomics Stage 2 projects undertaken, including the marker sets used – Angiosperms353 (Johnson et al. 2019); OZbaits (Waycott et al. 2021), and SaliBaits (A. Žerdoner Čalasan, unpublished) – current project status, and references for completed projects.

<table>
<thead>
<tr>
<th>Monophyletic group</th>
<th>Higher taxon</th>
<th>Marker set</th>
<th>Status</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acacia Mill.</td>
<td>Fabaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Alismatales</td>
<td></td>
<td>Angiosperms353, OZbaits</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Boronieae</td>
<td>Rutaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Chamaeleaucieae</td>
<td>Myrtaceae</td>
<td>Angiosperms353, OZbaits</td>
<td>Complete</td>
<td>Nge et al. (submitted)</td>
</tr>
<tr>
<td>Drosera L.</td>
<td>Droseraceae</td>
<td>Angiosperms353, OZbaits</td>
<td>Complete</td>
<td>Williamson et al. (submitted)</td>
</tr>
<tr>
<td>Ericaceae</td>
<td>Ericales</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Hibbertia Andrews</td>
<td>Dilleniaceae</td>
<td>Angiosperms353, OZbaits</td>
<td>Complete</td>
<td>Hammer et al. (submitted)</td>
</tr>
<tr>
<td>Lasiopetaleae</td>
<td>Malvaceae</td>
<td>Angiosperms353, OZbaits</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Mirbelieae</td>
<td>Fabaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Persoonia Sm.</td>
<td>Proteaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Santalaceae</td>
<td>Santalales</td>
<td>Angiosperms353, OZbaits</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Styliidaeae</td>
<td>Asterales</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Tecticornia Hook.f. / Salicornia L.</td>
<td>Chenopodiaceae</td>
<td>Angiosperms353, SaliBaits</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Teucrium L.</td>
<td>Lamiaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Lazarum A.Hay (as syn. Typhonium Schott)</td>
<td>Araceae</td>
<td>Angiosperms353, OZbaits</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
<tr>
<td>Xanthorrhoea Sol. ex Sm.</td>
<td>Asphodelaceae</td>
<td>Angiosperms353</td>
<td>Analysis underway</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Results

The Bioinformatics Working Group generated several key bioinformatics resources including: 1) a set of recommended approaches and software to use for preparation and analyses of target enrichment data for consortium-scale phylogenomics projects (link under GAP Data and Resources Availability section below); 2) updated scripts for the recommended software (Johnson et al. 2016, Yang and Smith 2014) to provide novel options, remove bugs, and seamlessly use the outputs of the former as inputs for the latter; 3) and a containerised analysis workflow for HybPiper 2.0 and ParaGone to overcome challenges with software compatibility issues and capability that are often encountered in herbaria that are not associated with universities or other research institutions, unfamiliarity with relevant programming languages, and the complexity involved in running numerous analysis steps required to use the recommended software (Jackson et al. 2023); 4) Python packages of HybPiper 2.0 and Paragone (links under GAP Data and Resources Availability section below). Additionally, work on AAToL led to two methods/resources studies. The first study, undertaken in partnership with PAFTOL researchers, developed a new target file – 'mega353' – for recovery of sequences of Angiosperms353 loci from target capture data sets (McLay et al. 2021). Use of this new target file, rather than the original target file of Johnson et al. (2019), can substantially increase the percentage of on-target reads, locus recovery at 75% length and the total length of the concatenated loci compared to the default Angiosperms353 file. The second study explored conflict in a range of AAToL Stage 2 datasets and developed guidelines for researchers to detect, characterise, manage and interpret conflict in target capture datasets (Joyce et al. 2024).

Together, the pilot and Stage 1 of the AAToL generated Angiosperms353 sequence data for 2,204 samples representing 2,019 genera. Of these, sequence data for 650 samples were provided by the PAFTOL project, and data for 75 samples were sourced from genomic repositories such as NCBI-SRA. A total of 2001 samples passed quality control, which has resulted in a draft phylogenomic tree including 1927 (90%) of 2144 native Australian angiosperm genera. Sequence data are available to the GAP consortium through the Bioplatforms Australia Data Portal. In a reciprocal data sharing arrangement with PAFTOL, data for 770 samples sequenced through the GAP Framework Initiative were made publicly accessible via the Kew Tree of Life Explorer (treeoflife.kew.org; Baker et al. 2022), which currently provides Angiosperms353 data for >10,000 species (data release 3.0, April 2023). AAToL contributed 763 samples (~8%) to a comprehensive analysis of the angiosperm tree of life including nearly 8000 (~60%) angiosperm genera which revealed new insights into angiosperm origins and diversification (Zuntini et al. 2024). The data have also contributed to several other publications produced by PAFTOL collaborators (e.g. Joyce et al. 2023, Elliot et al. 2024, Grass Phylogeny Working Group III 2024, Helmstetter et al. 2024, Pérez-Escobar et al. 2024). Samples released though the Kew Tree of Life Explorer are also publicly available from the Sequence Read Archive (European Nucleotide Archive – ENA – and National Centre for Biotechnology Information – NCBI; project PRJEB49212). The remaining data from Stage 1 of the AAToL will be made publicly available when the comprehensive phylogenetic analysis is completed and published.

The partnership with PAFTOL has greatly helped achieve GAP aims, most notably by (1) increasing the rate of progress toward completion of the AAToL; (2) expanding the professional networks of Australian scientists, particularly early to mid-career researchers, through collaboration with a global team of leaders in plant phylogenomics; and (3) maximising the potential for future re-use of GAP data through adoption of a universal target capture bait set and contribution to global datasets (Baker et al. 2022) and analyses (e.g. Zuntini et al. 2024). Taken together these outcomes substantially advance the capacity for genomics data generation and use by the Australian herbarium community and enable a step change in our understanding of the relationships and evolution of Australia's flora. In turn, GAP's engagement has significantly helped PAFTOL to broaden its reach, and has functioned as a model for the collaborative, open data mindset that PAFTOL aims to foster.

The Stage 2 of the AAToL generated Angiosperms353 sequence data for 10 species complexes (Table 2). In addition, sequence data was generated using OZbaits (Waycott et al. 2021) or SaliBaits (A. Žerdoner Čalasan pers. comm.) kits for several groups in which those kits were known, or expected, to outperform...
Angiosperms (Table 2). Data analysis, including an assessment of the utility of each bait set, and preparation of outputs will be reported in manuscripts focused on each taxonomic group. Papers on Chamelaucieae, Drosera and Hibbertia Andrews have been submitted. Several other studies utilising AAToL Stage 1 and Stage 2 data have also been published (e.g. on Aglaia Lour., Cooper et al. 2023; Pagonolepis Steetz, Schmidt-Lebuhn 2022; and Senecio L., Schmidt-Lebuhn et al. 2022).

The data and learnings from Stages 1 and 2 have enabled the community to imagine and commence planning for a continuation of the project beyond GAP: AAToL Stage 3. This stage 3 aims to resolve the AAToL to species level, complete for >95% of accepted Australian species. Avenues for resourcing this large undertaking (>18,000 species), and approaches for its completion are currently being explored.

Activity 3: Conservation Genomics

The Conservation Genomics stream aimed to provide genomic information to support conservation of the Australian flora. Conservation genomics covers a range of activities according to the management questions being addressed, including resolution of closely related taxa and species complexes to identify conservation dependent taxa; conservation units based on evolutionary lineages; genomic diversity to inform conservation interventions (e.g. augmentation, germplasm capture, translocation, restoration); hybridity and hybrid origin of species; mating system to understand level and pattern of outcrossing; and genotype-environment association analysis to identify signals of adaptation across climate gradients.

Of these types of studies, those aiming to resolve species complexes with suspected conservation-dependent species were deemed the most suitable for support through the GAP Framework Initiative. The Conservation Genomics Working Group was established to identify species complexes most in need of resolution due to conservation dependent units and taxonomic complexity, and to engage the research community in identifying groups where relevant taxonomic expertise could take advantage of the genomic input/data. The Working Group decided to take a specific molecular approach (double digest RADseq - ddRAD; Peterson et al. 2012) to both create interoperable data sets and develop capability for population-level data analyses. Projects to be supported were identified using the following criteria:

1. The species complex was expected to contain evolutionary units of conservation concern, and genomic data were required to inform taxonomic resolution.
2. Taxonomic expertise was available to utilise the genomic information.
3. Genomics expertise in data analysis and interpretation was available.
4. Samples were available as herbarium specimens, or fresh/dried material, and funding/resourcing to obtain samples was available.

A Request for Partnership elicited 15 responses involving 25 institutions, including seven state and institutional herbaria (Table 3). The Steering Committee was keen to stimulate collaboration among partners and the taxonomic and genomics communities and assisted by linking up collaborators in cases where proposals lacked key capability on the team (e.g. genomic data analysis expertise). Some projects did not proceed due to issues with DNA sample quality, particularly for projects that exclusively utilised herbarium samples. For five projects, the same samples were also sequenced using the Angiosperms353 bait kit to evaluate the comparative performance of a target capture approach for the resolution of species boundaries within species complexes.
Table 3. List of the GAP Conservation Genomics projects undertaken, including project status and the availability of target capture data (Angiosperms353) in addition to ddRAD data.

<table>
<thead>
<tr>
<th>Order</th>
<th>Family</th>
<th>Species complex</th>
<th>Status</th>
<th>Data generated</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asparagales</td>
<td>Orchidaceae</td>
<td><em>Paracaleana gracilicordata</em> Hopper &amp; A.P.Br.</td>
<td>Analysis underway</td>
<td>ddRAD</td>
<td>n/a</td>
</tr>
<tr>
<td>Asparagales</td>
<td>Orchidaceae</td>
<td><em>Thelymitra variegata</em> (Lindl.) F.Muell.</td>
<td>Project withdrawn</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Asterales</td>
<td>Asteraceae</td>
<td><em>Olearia ramulosa</em> (Labill.) Benth.</td>
<td>Analysis underway</td>
<td>ddRAD</td>
<td>n/a</td>
</tr>
<tr>
<td>Ericales</td>
<td>Ericaceae</td>
<td><em>Melichrus</em> R.Br.</td>
<td>Ended due to insufficient DNA sample quality. DArTseq (Kilian et al. 2012) and target capture (Angiosperms 353) data were subsequently generated for the same samples (H. Kennedy unpubl.).</td>
<td>DArTseq, Angiosperms353</td>
<td>n/a</td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td><em>Cassia</em> L.</td>
<td>Analysis underway</td>
<td>ddRAD, Angiosperms353</td>
<td>n/a</td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td><em>Gompholobium</em> Sm.</td>
<td>ddRAD completed; Angiosperms353 analysis underway</td>
<td>ddRAD, Angiosperms353</td>
<td>Simmons et al. (submitted)</td>
</tr>
<tr>
<td>Fagales</td>
<td>Casuarinaceae</td>
<td><em>Allocasuarina</em> L.A.S.Johnson</td>
<td>Ended due to insufficient DNA quality</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Liliales</td>
<td>Colchicaceae</td>
<td><em>Wurmbea dioica</em> subsp. <em>alba</em> T.D.Mcfar.</td>
<td>Analysis underway</td>
<td>ddRAD, Angiosperms353</td>
<td>n/a</td>
</tr>
<tr>
<td>Poales</td>
<td>Cyperaceae</td>
<td><em>Lepidosperma fimbriatum</em> Nees</td>
<td>Ended due to insufficient DNA quality</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Poales</td>
<td>Cyperaceae</td>
<td><em>Lepidosperma laterale</em> R.Br.</td>
<td>Ended due to insufficient DNA quality</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Proteales</td>
<td>Proteaceae</td>
<td><em>Isopogon buxifolius</em> R.Br.</td>
<td>ddRAD completed; ddRAD completed; Angiosperms353 analysis underway</td>
<td>ddRAD, Angiosperms353</td>
<td>Anderson et al. 2024</td>
</tr>
<tr>
<td>Proteales</td>
<td>Proteaceae</td>
<td><em>Synaphea stenoloba</em> A.S.George</td>
<td>Analysis underway</td>
<td>ddRAD</td>
<td>n/a</td>
</tr>
<tr>
<td>Sapindales</td>
<td>Rutaceae</td>
<td><em>Geleznowia verrucosa</em> Turcz.</td>
<td>Completed</td>
<td>ddRAD</td>
<td>Anderson et al. 2023</td>
</tr>
<tr>
<td>Sapindales</td>
<td>Rutaceae</td>
<td><em>Zieria</em> Sm.</td>
<td>Completed</td>
<td>ddRAD</td>
<td>Orel et al. 2023a, b</td>
</tr>
<tr>
<td>Sapindales</td>
<td>Simaroubaceae</td>
<td><em>Samadera bidwillii</em> (Hook.f.) Oliv.</td>
<td>Analysis underway</td>
<td>ddRAD</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Results

Of the 15 projects proposed, 11 provided DNA of suitable quality for ddRAD sequencing whereas four projects, on the genera *Lepidosperma* Labill. (x 2), *Allocasuarina* L.A.S.Johnson and *Melichrus* R.Br., did not proceed (Table 3). For most projects, herbarium specimens did not yield DNA of usable quality for ddRAD analysis and new collections were required to provide fresh material. The workflow for these projects included the assembly of sequence libraries and SNP calling using an optimised ipyrad (Eaton and Overcast 2020) approach, followed by custom filtering to provide high quality SNP data sets for downstream analyses. These analyses employed a combination of phylogenetic and population genetic approaches to obtain information on the genomic variation within and among the sampled populations of each species complex. This included concatenated and coalescent phylogenetic analyses, as well as ordination, admixture and distance-based population genetic analyses.

For each of the species complexes evaluated, the genetic results were interpreted in conjunction with morphological data to refine taxonomic hypotheses of relationships and to identify undescribed, sometimes cryptic, taxa or hybrid populations – results that will directly underpin conservation assessments and management actions. Further, the results provided insight into the morphological complexity within each group and the informative characters for species identification. Results for projects on *Geleznowia* Turcz., *Isopogon* R.Br. ex Knight and *Zieria* Sm. are already published (Anderson et al. 2023, 2024; Orel et al. 2023a, b).

Activity 4: Training

The GAP Framework Initiative aimed to build genomic capacity across Australian botanic gardens and herbaria and provide tools to enable genomic data to be used to identify, classify and conserve biodiversity. Training and capacity building across the GAP consortium has been an inherent and integral part of the development and deployment of workflows undertaken at all stages of the reference genomes, phylogenomics and conservation genomics projects. This included sampling of curated collections for genomic analysis, sample metadata curation, wet lab sample preparation for genomic analysis, bioinformatic analysis of genomic data, and preparation of manuscripts. The Wet Lab and Bioinformatics Working Groups reviewed current methods and available resources to develop best practice approaches and protocols, tools and training resources for GAP activities. Resources were also developed to train researchers in the use of the containerised bioinformatic pipelines (see below under Gap Data and Resources Availability).

Results

A webinar and workshop series for the analysis of target capture sequence data was delivered at the virtual Australasian Systematic Botany Society conference in July 2021 to 20 participants and the workshop materials and webinar recordings have been made publicly available (see below). Additionally, in response to key challenges identified in the early stages of GAP, resources were developed and made accessible to encourage and support best practice approaches for plant genomic initiatives and ongoing capacity building across the sector. These resources include: (a) protocols for the extraction of high molecular weight DNA for long read sequencing; b) analysis guides for target sequence capture projects; (c) containerised bioinformatic software pipelines for the analysis of target sequence capture data (Jackson et al. 2023); and (d) scripts for preparing target sequence data for upload to ENA. Links for the first three of the above public resources are provided in the GAP Data and Resources Availability section, below. The last will be made available in the near future.

In addition to the specific, measurable outputs detailed above, GAP has generated many ‘soft’ benefits for the community. A wide range of plant scientists, students and collections staff in herbaria, universities and botanic gardens – in total more than 210 people – were engaged in GAP activities. Mid-career and established researchers who have utilised legacy molecular approaches, such as Sanger sequencing or ISSR/SSR, for much of their careers have experienced a step change in their capacity to generate and
interpret genome scale datasets. Staff of smaller institutions that lack capacity for genetic research of any kind have enjoyed unprecedented opportunities to participate in large genomics projects. Postgraduate students and early career researchers already well versed in genomic analysis have been able to undertake specific training in areas of need and access large volumes of data that would have been difficult to finance otherwise. All participants have benefited from opportunities to engage and collaborate with leading researchers globally, and through those collaborations have developed a greater understanding of the applications of genomic technologies and data not only for evolution and systematics research, but also for collections development and management. These opportunities have undoubtedly led to higher-quality and more impactful research outputs, and expanded professional networks, both of which are likely to have a positive influence on careers.

GAP Outcomes

The GAP Framework Initiative has delivered (a) extended and strengthened collaborative networks, across local, national and international scales; (b) increased genomics resources and data; (c) enhanced community genomics capability; (d) made progress on the national strategic plans, and e) improved understanding of the Australian flora (Fig. 1).

Extended and strengthened collaborative networks

The GAP Framework Initiative consortium model has extended and strengthened collaborative networks across the sector nationally and internationally. Researchers and technical staff from the plant collections, systematics, genomics and bioinformatics communities, were represented from the initial consultations undertaken to develop the consortium through to the assembly of the Steering Committee and advisory Working Groups, and the wider consortium involved in different projects within the initiative Communication within and between Working Groups has strengthened cross-institutional collaboration and awareness on collections curation, wet lab, and bioinformatics activities, and promoted coordination of approaches and activities to maximise efficiency and the potential for data sharing and re-use. Beyond Australia, GAP researchers have developed deep collaborations across the global plant systematics and genomics community, principally through involvement in the PAFTOL consortium. These collaborations have produced high impact outcomes (e.g. McLay et al. 2021; Zuntini et al. 2024) and have sparked a large and diverse range of new projects with researchers worldwide (e.g. Elliott et al. 2024; Helmstetter et al. 2024). In turn, these relationships have influenced global thinking and practice, for example, through the endorsement and uptake of universal target capture bait kits.

Increased genomic resources and data

To date, the GAP framework initiative has generated reference genome sequence data for 34 species, including the first genomic resources for nine families and four orders, target sequence capture datasets representing 87% of Australia’s native flowering plant genera, and reduced representation genomic datasets for 10 species complexes of conservation concern. These outputs have considerably increased taxonomic coverage of genomic resources for Australian plants. All raw data are immediately released to GAP consortium partners on the Bioplatforms Australia Data Portal (https://data.bioplatforms.com/) and made accessible on public genomic repositories Australia upon publication of data papers. By mandating that source material must be vouchered in an Australian herbarium, the GAP framework Initiative has enhanced curated collections through the extended specimen (Lendemer et al. 2020, Webster, 2017) concept by enriching physical collections with genomic data.

Enhanced community genomics capability

The GAP consortium includes curators, researchers, technicians and other staff from more than 30 institutions across Australia and internationally. Consortium members have benefited from this collaborative program of work through their involvement in genomic data generation, analysis, interpretation and dissemination. These include sampling and curating metadata from herbarium
collections, wet lab preparation, bioinformatics analysis, preparation of research outputs and participation in training workshops. These researchers constitute an upskilled workforce equipped with improved capacity to generate and leverage genomic resources to identify, classify and manage biodiversity, and to improve the management, development and use of collections. Some institutions have already implemented improvements to data and specimen management protocols based on GAP processes and workflows.

Progress on national strategic plans

The outputs from the GAP Framework Initiative directly contributed to progress against various national science goals, in particular the Decadal Plan for Taxonomy and Biosystematics 2018-2027 (Taxonomy Decadal Plan Working Group 2018) through Strategic action 1.3 “…build a comprehensive framework to understand the evolution of the Australian and New Zealand biota.”, and Australia’s Strategy for Nature through Objective 10: “Increase knowledge about nature to make better decisions”, specifically progress measures 10A (Explicit science and knowledge programs to support effective management of biodiversity) and 10C (Systems capturing data on the diversity of Australian nature and how ecosystems function).

Improved understanding of the Australian flora

Australia’s biological heritage includes a diverse flora with a complex evolutionary history, widespread and restricted species, both long term natural isolation and more recent fragmentation, and both old and young lineages. GAP has helped improve upon Australia’s world-class infrastructure for understanding and managing this heritage by enhancing and enriching knowledge, as well as human and institutional resources. For example, the reference genomes are improving our understanding of genomic diversity, structure and evolution associated with major Australian plant radiations (e.g. Chen et al. 2022; McLay et al. 2022), the Australian Angiosperm Tree of Life will provide a common phylogenomic dataset for the flora and ultimately, the phylogenetic backbone for the Flora of Australia project and data portals such as the Atlas of Living Australia; and the conservation genomics projects are resolving the taxonomy and phylogeography of complex groups to enable species documentation and define units for conservation management (e.g. Anderson et al. 2023, 2024).

Lessons learned

The GAP project has not only built capacity and delivered data, but also revealed where the approaches taken were fit for purpose, and where they were not. For example, a key learning was the recognition that large collaborative consortium approaches such as GAP require dedicated human resourcing and robust governance structures. Critical to our success has been the provision of dedicated project management, community engagement and bioinformatics professionals, and steering and working groups overseeing and managing each of the major project domains. We learned the importance of converging on shared and scalable computational resources at the national level, while making use of common workflows tailored for GAP aims and use cases. Another key learning was understanding that a one-size fits all approach to source material/extraction protocols doesn’t work across the broad range of molecular approaches employed in this project. For example, while herbarium specimens generally yielded DNA suitable for target capture sequencing, they were generally unsuitable for the ddRAD analyses undertaken through conservation genomics projects, such that new, fresh collections were required. For the reference genomes projects most teams had access to adequate amounts of fresh tissue but still struggled to achieve the high quality, high molecular weight DNA extractions required for long read sequencing. Understanding and developing the most appropriate source material/extraction protocols for each approach was critical to success. Further investment in capacity building around DNA quality will be needed to further realise opportunities for genomics in collections science.

A final key learning was the value of large-scale collaborative consortia in rapidly increasing capacity and uptake of genomic approaches across a broad scope of research in genome assembly, phylogeny, systematics and conservation genomics. The project developed and strengthened collaborations and leveraged off the regional/local skills and specific taxonomic expertise across the group. The consortium
The approach used here provides a potential blueprint for similar initiatives in other countries, and a foundation for further development and application of genomics to plant research in Australia.

**GAP Data and Resources Availability**

Data generated by the GAP Framework Initiative adheres to FAIR principles – Findability, Accessibility, Interoperability, and Reusability (Wilkinson et al. 2016). These principles are achieved through: (a) immediate release of all data on the Bioplatforms Australia Data Portal to the GAP consortium, with an embargo period before public release; and (b) adopting genomic approaches in widespread global use that maximise the potential for reusability.

GAP data and other outputs are available as follows:

- Raw data – [https://data.bioplatforms.com/organization/about/bpa-plants](https://data.bioplatforms.com/organization/about/bpa-plants)
- Database of publicly available plant genome resources: [https://www.genomicsforaustralianplants.com/compilation-of-sequenced-plant-genomes/](https://www.genomicsforaustralianplants.com/compilation-of-sequenced-plant-genomes/)
- Reference genomes: *Acacia pycnantha* ([https://www.ebi.ac.uk/ena/browser/view/PRJNA752212](https://www.ebi.ac.uk/ena/browser/view/PRJNA752212)); *Telopea speciosissima* ([https://www.ebi.ac.uk/ena/browser/view/PRJNA712988](https://www.ebi.ac.uk/ena/browser/view/PRJNA712988))
- Reference genomes: Australian Reference Genome Atlas (ARGA) [https://arga.org.au](https://arga.org.au)
- Target capture sequences (AAToL): samples released through the Kew Tree of Life Explorer ([treeoflife.kew.org](https://treeoflife.kew.org)) are available from ENA (project PRJEB49212), and NBCI’s Short Read Archive (project PRJEB49212).

**Declaration of funding**

The Genomics for Australian Plants Framework Initiative consortium ([www.genomicsforaustralianplants.com/consortium/](http://www.genomicsforaustralianplants.com/consortium/)) is funded by Bioplatforms Australia (enabled by the National Collaborative Research Infrastructure Strategy – NCRIS), the Ian Potter Foundation, Royal Botanic Gardens Foundation (Victoria), Royal Botanic Gardens Victoria, the Royal Botanic Gardens and Domain Trust, the Council of Heads of Australasian Herbaria, CSIRO, Centre for Australian National Biodiversity Research and the Department of Biodiversity, Conservation and Attractions, Western Australia. The Plant and Fungal Trees of Life (PAFTOL) project at Royal Botanic Gardens Kew is funded by grants from the Calleva Foundation.

**Acknowledgements**

In addition to the funding partners listed above, we thank the leadership and staff of the many institutions that have made in-kind contributions including staff resources, access to collections, use of laboratories and other infrastructure. We specifically thank the following herbaria and their staff and collaborators who have collected, identified and curated the specimens used in this project: AD, BRI, CANB, CNS, DNA, MEL, MELU, NE, NSW, and PERTH. For the PAFTOL data used in the AAToL project, detailed acknowledgements are provided in Zuntini et al. (2024). Kor-Jent van Dijk, Rachael Fowler and Danielle Smith are thanked for
technical assistance, and Anna Monro is thanked for advice on Australian Plant Census records. We acknowledge the provision of computing and data resources provided by the Australian BioCommons Leadership Share (ABLeS) program (Gustafsson et al. 2023) which is co-funded by Bioplatforms Australia (enabled by NCRIS), the National Computational Infrastructure and Pawsey Supercomputing Centre. Additional compute was provided by various institutional facilities including the University of Melbourne and CSIRO.

Author contributions

Conceptualisation: M. Barrett, O. Berry, R. Binks, J. Bragg, L. Broadhurst, G. Brown, M. Byrne, D. Cantrill, D. Crayn, S. Ferguson, E. James, A. Jones, G. King, T. McLay, D. Murphy, K. Nargar, H. Sauquet, M. Waycott.


Funding Acquisition: L. Broadhurst, M. Byrne, D. Cantrill, D. Crayn.


Writing – original draft: D. Crayn with contributions from M. Byrne, D. Cantrill, L. Simpson.

References


Orel HK, Briggs JD, Duretto MF (2023b) Zieria nubicola (Rutaceae), a new and highly restricted species from New South Wales, Australia. Telopea 26, 169-174. Doi: 10.7751/telopea17501


Schmidt-Lebuhm AN, Cantrill DJ (2023) Genome size estimates for Genomics for Australian Plants sequencing projects. Muellera 41, 44-49.


