

Final report

Project title:

National Tree Genomics Program: Genomics Toolbox

Project leader:

Robert Henry

Delivery partner:

The University of Queensland

Report author/s:

Agnelo Furtado and Robert Henry

Project code:

AS17000

Date:

31 January 2024

Disclaimer:

Horticulture Innovation Australia Limited (Hort Innovation) makes no representations and expressly disclaims all warranties (to the extent permitted by law) about the accuracy, completeness, or currency of information in this final report.

Users of this final report should take independent action to confirm any information in this final report before relying on that information in any way.

Reliance on any information provided by Hort Innovation is entirely at your own risk. Hort Innovation is not (to the extent permitted by law) responsible for, and will not be liable for, any loss, damage, claim, expense, cost (including legal costs) or other liability arising in any way (including from Hort Innovation or any other person's negligence or otherwise) from your use or non-use of the final report or from reliance on information contained in the final report or that Hort Innovation provides to you by any other means.

Funding statement:

National Tree Genomics Program Genomics Toolbox is funded by the Hort Frontiers Advanced Production Systems Fund, part of the Hort Frontiers strategic partnership initiative developed by Hort Innovation, with co-investment from University of Queensland, BGI Australia, Bioplatforms Australia> and contributions from the Australian Government.

Publishing details:

Published and distributed by: Horticulture Innovation Australia Limited
ABN 71 602100149

Level 7
141 Walker Street
North Sydney NSW 2060

Telephone: (02) 8295 2300

www.horticulture.com.au

© Copyright 2024 Horticulture Innovation Australia Limited

Contents

Public summary	4
Keywords	4
Introduction	4
Methodology	5
Results and discussion	6
Outputs	11
Outcomes	12
Monitoring and evaluation	12
Recommendations	13
Refereed scientific publications	13
Intellectual property	15

Public summary

Genomics resources hold the key in reducing breeding cycles taking dozens of years, accelerating cultivar development in horticultural crops. The project generated genomic resources to support horticultural tree crop breeding and production for five crops (Almond, Avocado, Citrus, Macadamia and Mango). Improved reference genome sequences were generated, and resequencing defined the variation in important germplasm. Improved methods for the efficient generation of high-quality genome sequences were developed.

Almond: Sequencing of almond genotypes provided sequences for key Australian germplasm of interest to support breeding.

Avocado: A greatly improved reference genome sequence was produced for the variety “Hass”. All available Australian avocado varieties were sequenced.

Citrus: Reference genomes were produced for all 7 Australian native citrus species (including one undescribed species). Reference genomes were produced for 4 key Australian mandarin varieties.

Macadamia: Reference genomes were produced for all 4 species of Macadamia. Resequencing was applied to 365 genotypes representing the gene pool used in breeding and including all major commercial genotypes.

Mango: Three new reference genomes for mango were produced. Key mango germplasm including more than 400 genotypes were sequenced to support mango breeding.

The project delivered genomic resources that will support the breeding and production of all five crops. The methods developed will find application in other horticultural species.

Five Ph D students were trained in horticultural genomics.

All generated genomic resources are published and available for access.

Keywords

almond, avocado, citrus, mango, macadamia, genomics

Introduction

Horticulture is a highly valuable sector of the Australian agricultural industry. In 2013-14, the gross value of the Australian horticulture sector (excluding wine grapes) was \$8.73B, third in ranking behind the meat and grain industries (ABARES Agricultural commodities: September quarter 2015 - Statistics). Horticultural tree crops accounted for almost \$4.2B of revenue in 2016. (<http://horticulture.com.au/wp-content/uploads/2016/09/Horticulture-Statistics-Handbook-2015-16-Nuts.pdf>). Mango, macadamia, avocado, almond and citrus are the major tree crops of the Australian horticultural industry. Together, these five crops accounted for over 56% of horticulture tree crop revenue in 2016. Other key crops are apple, banana, and summer fruit, which together contributed a further 33% of revenue, making these eight crops accountable for almost 90% of the production by value.

While currently profitable, the horticulture tree industry faces numerous significant challenges that limit growth and fail to proactively address potential risk scenarios. As with many other crop industries, the fruit and nut sector operates within a highly dynamic environment that is heavily impacted by factors such as increasingly digital/automated technology, evolving consumer preference, higher associated costs and decreased labour availability, increasing environmental concerns, changing climate, the cost of energy, water and other resources, epidemics of new invasive insects and diseases and the need to support regional Australian communities. The horticulture tree industry is particularly susceptible to risk, as tree species have long breeding cycles and the generation and introduction of improved or recovered cultivars can require numerous breeding cycles, taking dozens of years. As it stands, the horticulture tree industry is generally reactive, even to identified, well-known risks. The long generation of tree crops makes it difficult for breeders and physiologists to proactively or rapidly develop new cultivars in response to pest and disease outbreaks, changing climate and evolving consumer preferences. Promising new advances in biotechnologies and genomics,

however, hold the key to accelerating cultivar development for horticulture trees. Genomic approaches offer opportunities to progress the efficiency of genetic improvement programs by developing more accurate predictions of the impact of underlying genetic variation on phenotypic performance. Improved understanding of the genetic control of biological processes influencing important traits can also be used to develop precision management practices. Importantly, new technical approaches can be applied to reduce the number and duration of breeding cycles in these trees. Studies in apple have demonstrated that elite seedlings can be propagated, and orchard planted as commercial cultivars after just 24 months using genomic prediction approaches - seven years earlier than through conventional breeding methods.

Australia's ability to be at the forefront of horticulture biotechnology is essential in ensuring that our horticulture industry remains profitable, productive, and as protected from potential risk as possible. A number of crops are important in the Australian horticulture industry; however, the long-lived nature of tree crops presents unique challenges for genetic improvement and horticulture management, including plant architecture, flowering, resistance to pests and disease, rootstocks and environmental (a biotic) stresses. The challenges faced are not insurmountable, but do require dedicated and concerted, interdisciplinary approaches that span genomics and genotype and phenotype prediction.

This project will establish a high-throughput genomics platform for mango, macadamia, avocado, almond and citrus. Tools will be delivered for use in genetic improvement and the study of the biology of these species for improved management. This project will provide a model for use in expanding the application of these technologies more widely across other horticulture crop species. Technology will find application in working with both rootstocks and scions for these species. The National Horticulture Tree Genomic Program aims to thoroughly investigate mango, macadamia, avocado, almond and citrus, with a large multinational, multidisciplinary team of world-renowned experts and industry bodies. As major synergies exist within the biological niche of tree species, much of the knowledge gained from investigation of one tree species in genetic improvement programs and management practices is transferable to other tree crops. Therefore, program results will be highly applicable to other species and concurrent research programs, providing maximum impact broadly across the horticulture tree sector.

Methodology

This project utilized next generation sequencing with long read technology to generate new reference genomes and short read technology to re-sequencing germplasm. Several Long-read technology platforms were compared to determine the best strategy (using one or combinations) for generating genome sequences (Murigneux et al., 2020). Following this, the amount of long-read sequence data required was evaluated to determine how much was required (Sharma et al., 2021) and new software tools were used to achieve de novo assembly (Sharma et al., 2022). These tools were then applied to establish the genetic relationships within the citrus, macadamia, and mango germplasm, to generate new reference genomes and to determine the sequence of relevant Australian germplasm.

Murigneux, V., et al and Henry, R.J. (2020) Comparison of long-read methods for sequencing and assembly of a plant genome. *GigaScience*, 9, giaa146 DOI: 10.1093/gigascience/giaa146).

Sharma, P., et al and Henry, R.J. (2021) Improvements in the sequencing and assembly of plant genomes. *GigaByte*, 2021, gigabyte24 DOI: 10.46471/gigabyte.24.

Sharma, P., et al and Henry, R.J. (2022) De novo chromosome level assembly of a plant genome from long read sequence data. *Plant J*, 109, 727-736 DOI: <https://doi.org/10.1111/tpj.15583>.

Results and discussion

Almond

Re-sequencing was used to determine the sequence of genotypes important for almond breeding in Australia. These data sets are made available to members working on other tool-box components of this project.

Avocado

Avocado genome:

Avocado (*Persea americana*) is a member of the magnoliids, an early branching lineage of angiosperms that has high value crop globally, having a highly nutritious profile including the unusual trait of storing fats in the fruit.

We generated a chromosome-level genome assembly for the commercial avocado cultivar Hass, which represents 80% of the world's avocado consumption (Nath et al., 2022). The DNA contigs produced from Pacific Biosciences HiFi reads were further assembled using a previously published version of the genome supported by a genetic map. The total assembly was 913 Mb with a contig N50 of 84 Mb. Contigs assigned to the 12 chromosomes represented 874 Mb and covered 98.8 % of benchmarked single-copy genes from embryophytes. Annotation of protein coding sequences identified 48,915 avocado genes of which 39,207 could be ascribed functions. The genome contained 62.6 % repeat elements. Specific biosynthetic pathways of interest in the genome were investigated. The analysis suggested that the predominant pathway of heptose biosynthesis in avocado may be through sedoheptulose 1,7 bisphosphate rather than via alternative routes. Endoglucanase genes were high in number, consistent with avocado using cellulase for fruit ripening. The avocado genome appeared to have a limited number of translocations between homeologous chromosomes, despite having undergone multiple genome duplication events. Proteome clustering with related species permitted identification of genes unique to avocado and other members of the Lauraceae family, as well as genes unique to species diverged near or prior to the divergence of monocots and eudicots.

This genome will support future advances in avocado biology and breeding/improvement.

Nath, O., Fletcher, S.J., Hayward, A., Shaw, L.M., Masouleh, A.K., Furtado, A., Henry, R.J. and Mitter, N. (2022) A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. *Horticulture Research*, 9, 157 DOI: 10.1093/hr/uhac157.

Citrus

Relationships between species:

Citrus is widely consumed across the world as a fruit crop. Despite many citrus species being extensively studied around the world, phylogenetic relationships among Australian native species remain unresolved.

This study has uncovered the phylogenetic relationships among six Australian native species, two domesticated citrus cultivars of commercial importance in Australia, and another thirteen accessions cultivated internationally based on complete, de novo assembled chloroplast genomes and 86 single copy nuclear genes (Nakandala et al., 2023).

The chloroplast and nuclear phylogenies were topologically different. The Australian species formed a monophyletic clade based on their nuclear genes. The nuclear phylogeny revealed a close relationship between *Citrus inodora* and *Citrus australasica*. These two species were distinct from the other four Australian limes that were more closely related to each other. *Citrus australasica* had a unique chloroplast which was distinct from all other Australian limes. Among the other Australian limes, *Citrus glauca* was the most distinct species based on nuclear genes, however, it had a similar chloroplast sequence to *C. australis*. The undetermined *Citrus* sp. was more closely related to *Citrus garrawayi*, indicating that it is a distinct form of *Citrus garrawayi*. *Citrus medica*, had a chloroplast similar to Australian species. However, the nuclear gene phylogeny analysis revealed that *C. medica* was more closely related to Asian citrus species. This study improves our understanding of phylogenetic relationships among Australian citrus species and confirms their

unique status within the genus since it formed a monophyletic clade which was clearly separated from the other non-Australian species.

Nakandala, U., Furtado, A., Smith, M.W., Williams, D.C. and Henry, R.J. (2023) Phylogenetic relationships among Australian native citrus species based upon complete chloroplast genomes and single copy nuclear genes. *Tropical Plants*, 2 DOI: 10.48130/TP-2023-0021.

Genome sequence of *Citrus australis*:

Recent advances in genome sequencing and assembly techniques have made it possible to achieve chromosome level reference genomes for citrus. Relatively few genomes have been anchored at the chromosome level and/or are haplotype phased, with the available genomes of varying accuracy and completeness.

We generated a phased high-quality chromosome level genome assembly for an Australian native citrus species; *Citrus australis* (round lime) using highly accurate PacBio HiFi long reads, complemented with Hi-C scaffolding (**Nakandala et al., 2023**).

Hifiasm with Hi-C integrated assembly resulted in a 331 Mb genome of *C. australis* with two haplotypes of nine pseudochromosomes with an N50 of 36.3 Mb and 98.8% genome assembly completeness (BUSCO). Repeat analysis showed that more than 50% of the genome contained interspersed repeats. Among them, LTR elements were the predominant type (21.0%), of which LTR Gypsy (9.8 %) and LTR copia (7.7 %) elements were the most abundant repeats. A total of 29,464 genes and 32,009 transcripts were identified in the genome. Of these, 28,222 CDS (25,753 genes) had BLAST hits and 21,401 CDS (75.8%) were annotated with at least one GO term. Citrus specific genes for antimicrobial peptides, defense, volatile compounds and acidity regulation were identified.

This chromosome scale, and haplotype resolved *C. australis* genome will facilitate the study of important genes for citrus breeding and will also allow the enhanced definition of the evolutionary relationships between wild and domesticated citrus species.

Nakandala, U., Masouleh, A.K., Smith, M.W., Furtado, A., Mason, P., Constantin, L. and Henry, R.J. (2023) Haplotype resolved chromosome level genome assembly of *Citrus australis* reveals disease resistance and other citrus specific genes. *Horticulture Research*, 10 DOI: 10.1093/hr/uhad058.

Genome sequence of *Citrus australasica*:

The finger lime (*Citrus australasica*), one of six Australian endemic citrus species shows a high natural phenotypic diversity and novel characteristics. The wide variation and unique horticultural features have made this lime an attractive candidate for domestication. Currently no haplotype resolved genome is available for this species.

We generated a high quality, haplotype-resolved reference genome for this species using PacBio HiFi and Hi-C sequencing (**Nakandala et al: Manuscript under review**).

Hifiasm assembly and SALSA scaffolding resulted in a collapsed genome size of 344.2 Mb and 321.1 Mb and 323.2 Mb size for the two haplotypes. The nine pseudochromosomes of the collapsed genome had an N50 of 35.2 Mb, 99.1% genome assembly completeness and 98.9% gene annotation completeness (BUSCO). A total of 41,304 genes were predicted in the nuclear genome. Comparison with *C. australis* revealed that 13,661 genes in pseudochromosomes were unique in *C. australasica*. These were mainly involved in plant-pathogen interactions, stress response, cellular metabolic and developmental processes, and signal transduction. The two genomes showed a syntenic arrangement at the chromosome level with large structural rearrangements in some chromosomes. Genetic variation among five *C. australasica* cultivars was analysed. Genes related to defence, synthesis of volatile compounds and red/yellow coloration were identified in the genome. A major expansion of genes encoding thylakoid curvature proteins was found in the *C. australasica* genome. The genome of *C. australasica* present in this study is of high quality and contiguity.

This genome helps deepen our understanding of citrus evolution and reveals disease resistance and quality related genes with potential to accelerate the genetic improvement of citrus.

Other Australian citrus species:

Australian wild limes holding the key to abiotic and biotic stress resilience are important genetic resources within the genus citrus.

We compared the chromosome-level genome assemblies of six Australian native limes, including four new chromosome level assemblies generated using PacBio HiFi and Hi-C sequencing (**Nakandala et al: Manuscript in advanced stage of preparation**).

The genomes were between 315 and 391 Mb with contig N50s from 29.5 to 35 Mb. Completeness of the assemblies was estimated to be from 98.4% to 99.3% and the annotations from 97.7% to 98.9% based upon BUSCO, confirming the high contiguity and completeness of the genomes. A high collinearity was observed among the genomes and the two haplotype assemblies for each species. Gene duplication and evolutionary analysis demonstrated that the Australian citrus have undergone only one ancient whole-genome triplication (γ -WGT event) event in their history. The highest number of species-specific and expanded gene families were found in *C. glauca* and they were primarily enriched in purine, thiamine metabolism, amino acids and aromatic amino acids metabolism which might help *C. glauca* to mitigate drought, salinity, and pathogen attacks in the drier environments in which this species is found. Unique genes related to terpene biosynthesis, glutathione metabolism, and toll-like receptors (TLRs) in *C. australasica*, and starch and sucrose metabolism genes in both *C. australis* and *C. australasica* might be important candidate genes for HLB tolerance in these species. Expanded gene families were not lineage specific, however, a greater number of genes related to plant-pathogen interactions, predominantly disease resistant protein (RPS2), was found in *C. australasica* and *C. australis*. These important gene families could be further explored to define their expression in response to stresses to facilitate their use in citrus improvement through molecular breeding.

Macadamia

The *Macadamia janseni* genome was used to develop the methods that were then applied to the other species. A total of 365 *Macadamia* genotypes were sequenced.

Genome sequence of *Macadamia janseni*:

Macadamia, a recently domesticated expanding nut crop in the tropical and subtropical regions of the world, is one of the most economically important genera in the diverse and widely adapted Proteaceae family. All four species of *Macadamia* are rare in the wild with the most recently discovered, *M. janseni*, being endangered. The *M. janseni* genome has been used as a model for testing sequencing methods using a wide range of long read sequencing techniques.

We have generated a chromosome level genome assembly of *Macadamia janseni* using a combination of Pacific Biosciences sequencing and Hi-C data sets (**Sharma et al., 2022**).

The *M. janseni* genome is comprised of 14 pseudo-molecules, with a N50 of 58 Mb and a total 758 Mb genome assembly size of which 56% is repetitive. Completeness assessment revealed that the assembly covered 96.9% of the conserved single copy genes. Annotation predicted 31,591 protein coding genes and allowed the characterization of genes encoding biosynthesis of cyanogenic glycosides, fatty acid metabolism and anti-microbial proteins. Re-sequencing of seven other genotypes confirmed low diversity and low heterozygosity within this endangered species. Important morphological characteristics of this species such as small tree size and high kernel recovery suggest that it is an important source of these traits for breeding. As a member of a small group of families that are sister to the core eudicots, this high-quality genome also provides a key resource for evolutionary and comparative genomics studies.

Sharma, P., Masouleh, A.K., Topp, B., Furtado, A. and Henry, R.J. (2022) De novo chromosome level assembly of a plant genome from long read sequence data. Plant J, 109, 727-736 DOI: <https://doi.org/10.1111/tbj.15583>.

Macadamia genomes – comparative genomics

Macadamia, a genus native to Eastern Australia, comprises four species, *Macadamia integrifolia*, *M. tetraphylla*, *M. ternifolia*, and *M. janseni*. Macadamia was recently domesticated largely from a limited gene pool of Hawaiian

germplasm and has become a commercially significant nut crop. Disease susceptibility and climate adaptability challenges, highlight the need for use of a wider range of genetic resources for macadamia production.

We generated high quality haploid resolved genome assemblies using HiFiasm to allow comparison of the genomes of the four species (**Sharma et al: Manuscript under review**).

Assembly sizes ranged from 735 Mb to 795 Mb and N50 from 53.7 Mb to 56 Mb, indicating high assembly continuity with most of the chromosomes covered telomere to telomere. Repeat analysis revealed that approximately 61% of the genomes were repetitive sequence. The BUSCO completeness scores ranged from 95.0% to 98.9%, confirming good coverage of the genomes. Gene prediction identified 37198 to 40534 genes. The species shared a common whole genome duplication event. Synteny analysis revealed a high conservation and similarity of the genome structure in all four species. Differences in the content of genes of fatty acid and cyanogenic glycoside biosynthesis were found between the species. An antimicrobial gene with a conserved cysteine motif was found in all four species.

The four genomes provide reference genomes for exploring genetic variation across the genus in wild and domesticated germplasm and to support plant breeding.

Diversity of wild *Macadamia*:

The genus *Macadamia* in the Proteaceae family includes four species native to Australia. Two of the four species *M. integrifolia* and *M. tetraphylla* have been recently utilized to generate domesticated macadamia varieties, grown for their edible nuts.

To explore diversity in macadamia genetic resources, a total of 166 wild genotypes, representing all four species, were sequenced, and subjected to phylogenetic analysis (**Manuscript: Manatunga et al., Submitted**).

The four species were clearly distinguished as four separate clades in a phylogenetic analysis of the nuclear genome (based upon concatenated nuclear genes). The two larger species (*M. integrifolia* and *M. tetraphylla*) formed a clade that had diverged from a clade including the smaller species (*M. ternifolia* and *M. janseni*). Greatest diversity in nuclear and chloroplast genomes was found in the more widely distributed *M. integrifolia* while the rare *M. janseni* showed little diversity. The chloroplast phylogeny (based upon whole chloroplast genome sequences) revealed a much more complex evolutionary history. Multiple chloroplast capture events have resulted in chloroplast genome clades including genotypes from different species. This suggests extensive reticulate evolution in *Macadamia* despite the emergence of the four distinct species that are supported by the analysis of their nuclear genomes. The chloroplast genomes showed strong associations with geographical distribution reflecting limited maternal gene flow in these species with large seeds. The nuclear genomes showed lesser geographical differences, probably reflecting longer distance movement of pollen.

This improved understanding of the distribution of diversity in *Macadamia* will aid conservation of these rare species now found in highly fragmented rainforest remnants.

Diversity of domesticated *Macadamia* and variant association analysis:

We have sequenced over 300 macadamia accessions covering all four species. A range of genomics analysis is underway to unravel the diversity with the domesticated accessions, heterozygosity and identifying SNP variants associated with phenotypic characters of interest to macadamia breeding. We expect to publish these results over several manuscripts.

Mango

Mango relationships:

The genus *Mangifera* (*Anacardiaceae*) includes 69 species of which *Mangifera indica* L. is the most important and primarily cultivated species for commercial mango production. Although the species are classified based on morphological descriptors, molecular evidence has confirmed the hybrid origin of two species suggesting the possibility that more of the species may be of hybrids origin.

To analyse evolutionary relationships within the genus, 13 samples representing 11 *Mangifera* species were sequenced and whole chloroplast genomes and 47 concatenated common single-copy nuclear genes sequences were assembled and used for phylogenetic analysis (Wijesundera et al: Manuscript submitted).

The chloroplast (Cp) genome size varied from 151,752 to 158,965 bp with *M. caesia* and *M. laurina* having the smallest and largest Cp genome respectively. Chloroplast genome annotation revealed 80 protein coding genes, 31 tRNA and four rRNA genes across all the species.

Comparative analysis of whole chloroplast genome sequence and nuclear gene-based phylogenies revealed topological conflicts suggesting chloroplast capture or cross hybridization. The chloroplast of *M. altissima*, *M. applanata*, *M. caloneura* and *M. lalijiwa* were similar to those of *M. indica* (99.9% sequence similarity). Their close phylogenetic relationship suggests a common ancestry and the likely cross hybridization of these wild species with *M. indica*. Similarly, topological incongruence observed for *M. casturi* and *M. zeylanica* suggests cross-hybridization and repeated backcrossing with *M. indica* in which the paternal parent was *M. indica*. This study provides improved knowledge of the phylogenetic relationships and evidence of gene flow among the different species suggesting that hybrids may be common within the genus *Mangifera*.

Genome sequence of *Mangifera indica* cv Irwin:

Improvements in long-read sequencing techniques have greatly accelerated plant genome sequencing. Current *de novo* assemblies are routinely achieved by assembling long-read sequence data into contigs that are assembled to chromosome level by chromatin mapping (Hi-C).

We report here a telomere-to-telomere chromosome-level genome of the mango cultivar Irwin, using only PacBio HiFi long reads. (Wijesundara et al: Manuscript submitted).

HiFi reads at high coverage (204X) resulted in the assembly of 17 chromosomes each as a single contig with telomeres at both ends. The remaining three chromosomes were represented each by two contigs with telomeres at one end and ribosomal repeats at the other end. Analysing contig ends allowed them to be paired and linked to generate the remaining three complete chromosomes, telomere-to-telomere. The assembled genome was 365 Mb with 100% completeness as assessed by BUSCO analysis. The haplotypes assembled from the same data demonstrated extensive structural differences.

This approach may be useful for assembling telomere-to-telomere and haplotype-resolved genomes for many other plants.

Mango genomes – comparative genomics

Chromosomal level genome assembly was generated for three mango cultivars cv., Kensington Pride, Lauria Lombok and Irwin. These genomes and their gene complement are being analysed to determine their genome structure and difference in their genes (Wijesundera et al, Manuscript under preparation).

Sequence database of full-length transcript variants of genes

A collection of transcript sequences representing expressing variants of all genes in mango can be an important resource to study gene expression. Full length sequencing of transcripts using PacBio Iso-sequencing was used to generate transcript data of genes expressing in leaf and root tissue. This database can also be used for the annotation of assembled genomes and to measure gene expression (Wijesundara et al: Manuscript in prep).

Outputs

Table 1. Output summary

Output	Description	Detail
A reference genome for avocado	Nuclear genome sequence of Avocado cultivar Hass published and freely available. All data freely available..	Nath, O. , Fletcher, S.J., Hayward, A., Shaw, L.M., Masouleh, A.K., Furtado, A., Henry, R.J. and Mitter, N. (2022) A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. Horticulture Research, 9, 157 DOI: 10.1093/hr/uhac157.
Reference genomes for citrus	Nuclear genome sequences of two Australian wild Citrus published and freely available. Another five to be published and to be freely available. All data freely available	Nakandala, et al (2023) Horticulture Research, 10 DOI: 10.1093/hr/uhad058. Nakandala, U. , et al (2023) Horticulture Research, 10 DOI: 10.1093/hr/uhad058. Manuscript under review: Comparative genomics study of all four macadamia genomes.
Reference genomes for mango	Three Nuclear genome sequences: Two <i>M. indica</i> accessions (Kensington Pride) and Irwin <i>M. Laurina</i> cv Lombok	Publications underway and all data sets will be freely available.
Reference genomes for macadamia	Nuclear genomes for four species: <i>M. jansanii</i> , <i>M. integrifolia</i> , <i>M. tetraphylla</i> and <i>M. ternifolia</i>	Sharma, P. , et al. (2022) De novo chromosome level assembly of a plant genome from long read sequence data. Plant J, 109, 727-736 DOI: https://doi.org/10.1111/tpj.15583 . Manuscript under review: Comparative genomics study of all four macadamia genomes.
Almond genome sequences	Sequence data sets of 10 Almond accessions generated and distributed researchers from other Tool-boxes	Data can be made available
Avocado genome sequences	Sequence data sets of 54 Avocado accessions generated and processed for variant analysis	Data can be made available
Macadamia genome sequences	Sequence data sets of over 300 macadamia accessions generated and processed for variant analysis	Data can be made available
Mango genome sequences	Sequence data for 10 wild accessions and four <i>M. indica</i> cultivars	Data can be made available

Outcomes

Table 2. Outcome summary

Outcome	Alignment to fund outcome, strategy and KPI	Description	Evidence
Reference genomes for use by plant breeders	Tools and Advances technologies: Plant Genetics: Next generation sequencing was used to obtain genome sequences of Avocado, Macadamia, Citrus and Mango accessions.	Results are to be published and all information including data sets are freely available for research.	Breeding program and project participants have access to the results and genome sequences. This data will feed into other programs for improved cropping systems and increased productivity
Genome sequences for use in plant variety identification	Tools and Advances technologies: Plant Genetics: Identification of variety specific markets will aid in plant breeding	Results are to be published and all information including data sets are freely available for research.	Species information has been amended in the Macadamia germplasm collection. Avocado: Data generated is made available for varietal identification.
Genomes sequences for use in gene discovery	Tools and Advances technologies: Plant Genetics: Identification of genes will aid in selection via plant breeding leading to increased productivity.	Results are to be published and all information including data sets are freely available for research.	Citrus: Important gene identified with function associated with biotic stress resistance.
Genome sequences for use in protection of intellectual property	Tools and Advances technologies: Plant Genetics: Identification and protection of varieties	Genotype specific variants can be made available or protected for use by the industry	Mango: sequence data was used to identify varietal-specific variants to confirm a variety.

Monitoring and evaluation

Table 3. Key Evaluation Questions

Key Evaluation Question	Project performance	Continuous improvement opportunities
Workshop at Project initiation: Workshop End Year and 2 between industry groups and researchers	Instructed and helped shape the project aims based on industry stakeholders needs. Subsequent workshops help understand outcomes, discoveries made and research and industry collaboration.	Additional workshops will help this process.

Gap analysis	Completed in Year 1 and continued through subsequent years. Instructed prioritizing generation of genomic resources.	Improved genomic resources using continuously improving technologies to match evolving needs of the industry.
Project management: Regular meetings: HortInnovation PM, Project CI's and industry stakeholders (researchers and breeders). Risk management plan	Revised milestones Table in Year 1 completed and approved. All milestones completed. Delayed start of PhD students (Covid-related) was successfully managed via the Risk management plan.	This process has been helpful.

Recommendations

- Molecular breeding should be adopted in all five species using the platform delivered in the project
- Molecular techniques are available for the identification of varieties of these five species and can be used to protect new varieties and to confirm the identity of samples in cases of uncertain identity.

Refereed scientific publications

Journal article

- Nakandala U**, Furtado A, Smith MW, Williams D, Henry RJ (2023) Phylogenetic relationships among Australian native citrus species based upon complete chloroplast genomes and single copy nuclear genes. Tropical Plants doi: [10.48130/TP-2023-0021](https://doi.org/10.48130/TP-2023-0021)
- Nakandala U**, Masouleh AK, Smith MW, Furtado A, Mason P, Constantin L, Henry RJ (2023) Haplotype resolved chromosome level genome assembly of *Citrus australis* reveals disease resistance and other citrus specific genes. Horticulture Research doi.org/10.1093/hr/uhad058
- Henry RJ** (2023) Genomic analysis of new food options from the Australian flora. Sustainable Food Technology DOI: 10.1039/D3FB00001J
- Nath O**, Fletcher SJ, Hayward A, Shaw LM, Masouleh AK, Furtado A, Henry RJ, Mitter N (2022) A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. Horticulture Research <https://doi.org/10.1093/hr/uhac157>
- Nath O**, Fletcher S, Hayward A, Agarwal R, Furtado A, Henry RJ, Mitter N (2022) A comprehensive high-quality DNA and RNA extraction protocol for a range of cultivars and tissue types of the woody crop avocado. Plants 11 (3), 242
- Sharma P**, Masouleh AK, Topp B, Furtado A and Henry RJ (2021) De novo chromosome level assembly of a plant genome from long read sequence data. The Plant Journal <https://doi.org/10.1111/tpj.15583>
- Sharma P**, Murigneux V, Haimovitz J, Nock CJ, Masouleh AK, Topp B, Furtado A, Henry RJ (2021) The genome of the endangered *Macadamia janseni* displays little diversity but represents an important genetic resource for plant breeding. Plant Direct 5(12), e364. <https://doi.org/10.1002/pld3.364>
- Sharma P**, Aldossary O, Alsubaie B, Al-Mssallem I, Nath O, Mitter N, Rodrigues Alves Margarido G, Topp B, Murigneux V, Kharabian Masouleh A, Furtado A, Henry RJ (2021) Improvements in the sequencing and assembly of plant genomes. Gigabyte doi: 10.46471/gigabyte.24
- Sharma P**; Al-Dossary O; Alsubaie B; Al-Mssallem I; Nath O; Mitter N; Rodrigues Alves Margarido G; Topp B; Murigneux V; Masouleh AK; Furtado A; Henry RJ (2021): Supporting data for "Improvements in the Sequencing and Assembly of Plant Genomes" GigaScience Database. <http://dx.doi.org/10.5524/100906>

- Murigneux V**, Rai SK, Furtado A, Bruxner TJC, Tian W, Ye Q, Wei H, Yang B, Harliwong I, Anderson E, Mao Q, Drmanac R, Wang O, Peters BA, Xu M, Wu P, Topp B, Coin LJM, Henry RJ (2020) Comparison of long read methods for sequencing and assembly of a plant genome. *GigaScience* 9, giaa146 <https://doi.org/10.1093/gigascience/giaa146>
- Nock CJ**, Baten A, Mauleon R, Langdon KS, Topp B, Hardner C, Furtado A, Henry RJ and King GJ (2020) Chromosome-scale assembly and annotation of the macadamia genome (*Macadamia integrifolia* HAES 741). *G3: Genes Genomes Genetics* 10, 3497-3504. doi: 10.1534/g3.120.401326

Chapters in a book or paper in conference proceedings

- Okemo P**, Wijesundara U, Nakandala U, Ananda GKS, Vanambathina P, Hasan S, Abdulla M, Priyanka Sharma, Manatunga S, Pazhany A, Masouleh AK, Nath O, Mitter N, Furtado A, Henry R J (2022) The use of genome sequencing to improve crops for tropical agriculture. In: Next generation sequencing and agriculture, CABI, Bayer P, Edwards, D Eds. 9 92-138. DOI: 10.1079/9781789247848.0004
- Shaw L**, Onkar Nath, Fletcher S, Henry R, Mitter N, Hayward A (2021) Avocado Transcriptomic Resources. In: Cifuentes, A. (Ed.), Comprehensive Foodomics, Volume 1 Elsevier, pp 544-557.
- Manatunga SL**, Furtado A, Topp B, Kharabian Masouleh A, Henry R (2024) Genome Diversity of Macadamia. Plant and Animal Genome Conference San Diego
- Henry RJ** (2024) Genomics of crop wild relatives from the Australian flora. Plant and Animal Genome San Diego
- Henry RJ** (2023) De novo genome sequences for agricultural biotechnology. Agricultural Biotechnology Present and Future SABC Symposium
- Henry RJ** (2023) Genomics of Australian plants to support diversification of agriculture. Plant and Animal Genome Australia Perth
- Wijesundara U**, Masouleh AK, Furtado A, Dillon NL, Henry RJ (2023) A highly accurate chromosome-scale genome assembly for mango (*Mangifera indica* L. cv Irwin). Plant and Animal Genome Australia Perth
- Henry RJ** (2023) Chromosome level assembly of plant genomes. Plant and Animal Genome Australia Perth
- Wijesundara WWMUK**, Furtado A, Dillon N, Kharabian Masouleh A, Henry R (2023) Determination of species identity and genetic diversity in genus *Mangifera* based on chloroplast and nuclear genome sequences. International Mango Symposium, Spain
- Henry R J** (2023) Genomics of Australian crop wild relatives, a key resource for global food security. International Congress off genetics, Melbourne
- Henry R**, Furtado F, Masouleh A, Abdullah M, Morrison S, Upuli Nakandala U, Shama P, Fang J, Mason P, Chandora R, Okemo P (2023) Sequencing the Genomes of Australian Crop Wild Relatives. International Congress of Genetics, Melbourne
- Henry RJ** (2023) Genomics of tropical plants 1st International Conference on Tropical Plants Sanya China.
- Henry** (2023) Using genomics to capture plant biodiversity for agriculture. International academic Forum of Beijing Academy of Agricultural and Forestry Sciences
- Henry RJ** (2023) Reference genomes to support crop breeding. Plant and Animal Genome Conference, San Diego
- Sharma P**, Kharabian Masouleh A, Topp B, Furtado A, Henry RJ (2023) Comparison of the genomes of the *Macadamia* species. Plant and Animal Genome Conference, San Diego
- Wijesundara WWMUK**, Furtado A, Dillon N, Kharabian Masouleh A, Henry R (2023) Chloroplast and nuclear genome sequence based phylogenetic relationships, species distinction, and genetic diversity in the genus *Mangifera*. Plant and Animal Genome Conference, San Diego
- Nakandala U**, Furtado A, Smith MW, Kharabian-Masouleh A, Williams D, Henry RJ (2023) Chromosome level assembly and annotation of Australian native and domesticated citrus species. Plant and Animal Genome Conference, San Diego
- Manatunga S**, Furtado A, Topp B, Kharabian Masouleh A, Henry R (2022) Sequencing Diversity of Macadamia Species. TropAg 2022
- Nakandala U**, Furtado A, Smith M, Kharabian-Masouleh1 A, Williams D, Henry R (2022) Phylogenetic relationships between Australian native citrus and commercial varieties. TropAg 2022

Sharma P, Kharabian Masouleh A, Topp B, Furtado A, Henry RJ (**2022**) Genomes of the four Macadamia species. TropAg 2022

Soni A, Henry R, Furtado A, Okemo P (**2022**) Studying the evolutionary significance of genome size diversity in the Proteaceae family. TropAg 2022

Wijesundara1 UWWM, Furtado A, Dillon N, Kharabian Masouleh A, Henry R (**2022**) Analysis of phylogenetic relationships in genus Mangifera using chloroplast and nuclear phylogeny. TropAg 2022

Henry RJ (**2022**) Sequencing macadamia genomes. PAG Asia

Henry RJ (**2022**) National Tree Genomics Program. National Horticulture Research Network Developing a national approach to examining Australian horticultural germplasm collections for germplasm to breed varieties for a changing climate. August 30 & 31 2022, Waite Campus, University of Adelaide

Henry RJ (**2022**) Improving reference genomes to support genomic breeding. PAG Asia

Henry R J (**2022**) Sequencing and Assembly of diverse plant genomes. Genomics of Animals and Plants Asia, Dovetail 20 April Singapore

Henry RJ (**2022**) The future of food- insights for the agricultural and food industries. The Australian and New Zealand Society of Occupational Medicine Inc. Brisbane 21 March.

Henry R J (**2022**) Sequencing and assembly of plant genomes: advances in methods and applications. 43rd Annual Meeting of Plant Tissue Culture Association (India) & International Symposium on Advances in Plant Biotechnology and Nutritional Security (APBNS-2022) New Delhi Feb 18th

Henry RJ (**2021**) Genomics of wild and domesticated plant diversity for crop improvement. Joint Genome Institute Annual Research Meeting September San Francisco 30 August 3 to 1 September

Sharma P, Furtado A, Nock C, Topp B and Henry R (**2019**) Comparative genomics of macadamia species. TropAg 2019

Intellectual property

No project IP or commercialisation to report