Hort Innovation

Final Report

Genetic diversity and population structure of wild and domesticated macadamia

Project leader:

Dr Catherine Nock

Delivery partner:

Southern Cross University

Project code:

MC18004

Project:

Genetic diversity and population structure of wild and domesticated macadamia (MC18004)

Disclaimer:

Funding statement:

This project has been funded by Hort Innovation, using the macadamia research and development levy and contributions from the Australian Government. Hort Innovation is the grower-owned, not-for-profit research and development corporation for Australian horticulture.

Publishing details:

ISBN 978-0-7341-4813-1

Published and distributed by: Hort Innovation

Level 7
141 Walker Street
North Sydney NSW 2060

Telephone: (02) 8295 2300

www.horticulture.com.au

Content

Summary

Macadamia is a major contributor to the Australian horticultural industry and is the first Australian native plant to have been developed as an international food crop. Wild populations, found only in subtropical eastern Australia, are an irreplaceable source of variation that will be valuable in the development of improved varieties for the global macadamia industry.

This project aimed to advance understanding of the population structure and genetic diversity in macadamia, to support the long-term conservation of genetic and morphological diversity for the benefit of the macadamia industry. It involved sampling, genotyping and analysis of over 1000 accessions including wild and planted trees, predominant cultivars, new levy-funded cultivars and germplasm from the Australian and Hawaiian breeding programs.

Genotype data was used to identify individual trees as *M. integrifolia* (427), *M. tetraphylla* (314)*, M. ternifolia* (40)*, M. jansenii* (182) and interspecific hybrids (68). Identification of each of the four species, interspecific hybrids and individuals was achieved with 15 macadamia-specific SSR genetic markers. Select subsets of these markers are useful for cultivar identification and paternity analysis.

Distinct genetic clusters were identified in *M. integrifolia* and *M. tetraphylla.* Results indicate that genetic variation is geographically structured in *M. integrifolia* and includes a northern group with two clusters: (1) Mt Bauple and (2) Gympie region from Mary Creek to Amamoor; and a southern group with three clusters: (3) north of the Brisbane River from Samford to Villeneuve, (4) Brisbane from Holland Park (Sankey's Scrub) to Mt Cotton, (5) Mt Cotton to the Gold Coast hinterland. In *M. tetraphylla,* there was a weaker correlation between genetic and geographic distance, and genetic clusters included trees from geographically distant sites. However, distinct clusters were identified at (1) Duroby, (2) Mt Warning, Uki, Lismore, Nimbin, Guanaba (3) Lennox, Tyalgum (4) Victoria Park, Mt Glorious, Numulgi (5) Meerschaum Vale, Victoria Park and Coolgardie.

The Brisbane cluster included some historically important trees including the Walter Hill tree at the Brisbane Botanical Gardens in 1858, one planted at the University of California Berkely USA in 1879 and a tree at John MacArthur's Menangle estate thought to be over 150 years old that was found through the citizen science project, the Wild Macadamia Hunt.

There was evidence of extensive admixture between regions as a result of natural or human-mediated dispersal. The proposed mechanisms for natural dispersal include bee pollination, and seed dispersal by small mammals, gravity and water. Natural dispersal over large distances (>3 km) is expected to be limited. However, connectivity between habitat fragments may be supported by intermediate trees and populations. Admixture was greatest within the southern range of *M. integrifolia* and across the much of the distributional range of *M. tetraphylla*. Some sites showed little evidence of admixture and may represent original wild populations. These sites and individual trees assigning >80% to each of the genetic clusters could be considered when prioritising populations for conservation, and selecting trees for inclusion in *ex situ* collections.

Results suggest that most Hawaiian cultivars are *M. integrifolia* and closely related to one another. The Australian cultivars are more genetically diverse and include many interspecific *M. integrifolia x M. tetraphylla* hybrids. Interestingly, a number of recently released cultivars are also hybrids suggesting that interspecific hybridisation could be important in breeding new, improved varieties.

The conservation of key natural populations of all *Macadamia* species will be important for future improvement of this native Australian nut crop. Most cultivars are derived from the northern *M. integrifolia* group at Mt Bauple and the Gympie region. The southern *M. integrifolia* group and wild populations of *M. tetraphylla*, *M. ternifolia* and *M. jansenii* represent novel sources of genetic and morphological diversity that could drive future productivity.

Keywords

macadamia; population structure; conservation; genetics; breeding

Introduction

Macadamia is the first Australian native plant to have been developed as an international food crop. Wild populations are found only in subtropical eastern Australia. Worldwide, there is increasing recognition of the importance of conserving crop wild relatives for future food security (Bronzynska et al. 2015).

The genus *Macadamia* includes four species with a discontinuous distribution of over 700 km in northeast NSW and southeast Queensland. All species are listed as threatened and their habitat, the lowland rainforest of subtropical Australia, is listed as a critically endangered ecosystem. For over 30 million years, macadamia have evolved and adapted to their environments generating considerable genetic and morphological variation across the genus (Mast et al. 2009).

Until recently, limited genetic data for macadamia were available. Southern Cross University led the Australian effort in collaboration with University of Queensland to sequence the HAES 741 genome (Nock et al. 2014; Nock et al. 2020) and develop genetic linkage maps for macadamia (Langdon et al. 2020). These genomic resources have contributed to other Hort Innovation funded projects including MC14000 *Macadamia Second Generation Breeding and Conservation*, MC15007 *Still wild about macadamias – conserving a national icon* and AS17000 *National Tree Genomics Program*.

Genetic markers developed at Southern Cross University have been used to support the Australian macadamia industry, including the DNA authentication of new releases. These SSR markers have also been used for DNA paternity testing (MC09021 *Macadamia Breeding and Conservation*, Langdon et al. 2019, Kamper et al. 2021) and are being used to support breeding in South Africa and China (Dr Heng Lui, Dr Gerda Fourie *pers. comm.*). Genotyping of wild and domesticated germplasm is crucial to identify population structure and assist in the long-term conservation of maximum genetic diversity for future crop improvement.

In 1996, the National Macadamia Germplasm Collection was established with clones of trees sampled from presumed wild populations the species that have been used in cultivation: *M. integrifolia*, *M. tetraphylla* and *M. ternifolia*. Since then, many new wild populations and old remnant trees have been located following extensive survey work conducted as part of the Hort Innovation project MC14000 *Still Wild about Macadamia* and Healthy Land and Water Qld's citizen science project *Wild Macadamia Hunt,* and by the Macadamia Conservation Committee, Southern Cross University and University of Sunshine Coast.

Recent research indicates that genetic diversity in *M. integrifolia* is partitioned into five distinct genetic-geographic groups with historical barriers to seed dispersal between regions. The wild origins of the Hawaiian cultivars were traced back to Mooloo and Mt Bauple in southeast Queensland and indicate that a bottleneck occurred during domestication. One genetic group was represented only by cultivated trees suggesting that the genetics of extinct wild populations in the Brisbane region may be preserved in trees planted in parks, gardens and backyards (Nock et al. 2019).

The main objective of this project was to define the distribution and structure of genetic diversity in wild, planted and domesticated macadamia. This will support the prioritisation of populations and trees for conservation, and the selection of novel variation for inclusion in future *ex situ* and breeding collections.

Methodology

Over 1000 macadamia accessions were sampled and genotyped for this project and DNA from a select group of accessions was genome resequencing of DNA from a subset of selected wild accessions. The methodology for sample collection, DNA extraction and SSR genotyping has been previously described (Nock et al. 2014, Langdon et al. 2019).

Sample collection

Leaf samples of 79 industry cultivars and breeding selections were sourced from reference collections and regional varietal trials in Australia including the Australian Macadamia Breeding Program, Hidden Valley Plantations and United States Department of Agriculture, Hawaii and University of Adelaide.

Samples were collected from the National Macadamia Germplasm Collection, and additional wild populations and old trees were located and sampled through a major effort involving a large group including the Macadamia Conservation Committee, Healthy Land and Water Qld, the Wild Macadamia Hunt, landholders,

growers, Southern Cross University, University of the Sunshine Coast, University of Queensland, NSW DPI and QDAF. This represents the largest collection of macadamia accessions ever genotyped (Table 1). Leaf material was dried on collection with approximately 10 times wet leaf weight of silica and sealed in zip-lock bags. Where possible, young leaf or shoot material was collected. Samples were stored at room temperature prior to DNA extraction.

Species	**Accessions**
M. integrifolia	427
M. tetraphylla	314
M. ternifolia	40
M. jansenii	182
Hybrids	68
Not genotyped	36
Total	**1067**
Cultivars	79

Table 1. Summary of macadamia accessions sampled.

DNA extraction

DNA was extracted using Qiagen DNeasy kits and methods optimised for recovery of high-quality DNA from silica-dried macadamia leaf. It is important to note that it is often not possible to collect young leaf or shoot material from wild trees. Our experience, based on DNA extraction from thousands of macadamia leaf samples, indicates that the high quality/quantity DNA required for genome resequencing and genotype by sequencing approaches such as DArT (minimum 50 ng/µl) is usually only recovered from fresh or young leaf/shoot. SSR genotyping is a robust method used in human forensics and agriculture that requires minimal DNA for genotyping (5 ng/µl) due to PCR amplification of the SSR loci.

Genotyping

SSR: Genetic short sequence repeat (SSR) markers were developed specifically for macadamia (Nock et al. 2014, Langdon et al. 2019). These have been applied for the Australian industry for cultivar identification to ensure that 'true to type' scionwood from newly released selections is used for nursery propagation. Subsets of 15 multi-allelic macadamia SSR markers can distinguish between cultivars, individuals, hybrids and species and are being utilised for DNA paternity analysis in orchards and gene flow studies in wild populations. Some of the benefits of SSR genotyping are that the primers and the methods are published, non-proprietary and transferrable between labs.

Whole genome resequencing: Collaboration with international colleagues, Professor Ray Ming, Dr Jishan Liu and others supported whole genome sequencing of high-quality DNA (>100 ng/ul) from a select group of wild accessions to determine the Australian origins of macadamia domestication. A manuscript on the genome of HAES 344 and the domestication origins of macadamia is currently under review and will be available on publication.

Genetic Analysis

A range of population genetic and bioinformatic methods was applied for analysis of SSR and resequencing datasets to determine the genetic diversity and population structure of *Macadamia* across the geographic distribution of the genus. Analyses conducted included population structure to determine major genetic groups in *Macadamia,* and within species. Relative diversity and relatedness to determine relationships between cultivars and potential source populations. Genetic diversity measures for each species and cultivars, pairwise relatedness, principal coordinate analysis and clone identification were performed in GenAlEx 6.503. STRUCTURE 2.3.4 and Structure Harvester were used for species assignment and population structure analysis.

Outputs

Project outputs include:

- A final report containing genotypic information of identified accessions and passport information, raw and analysed data illustrating genetic diversity and population structure of wild and domesticated Macadamia accessions, and a list of accessions that are of likely conservation value for future productivity (page 10-17).
- A non-technical summary of findings for dissemination to growers and industry stakeholders (page 4).
- News bulletin articles (Appendix 2 and 3)
- Refereed scientific articles and conference presentations (page 22)

Genetic diversity and population structure in Macadamia

Results and Discussion

Macadamia species

Population structure analysis of all accessions including wild and planted trees and cultivars identified four main genetic groups corresponding to the four species *M. integrifolia, M. tetraphylla, M. ternifolia* and *M. jansenii* as well as interspecific hybrids (Figure 1). A recent genetic study of macadamia using DArT markers identified a close relationship between the two small tree species *M. ternifolia* and *M. jansenii* which formed a single cluster (Mai et al. 2020). This may have been due to the small numbers of individuals available. We analysed the genotypes of 179 *M. jansenii* and 42 *M. ternifolia* and found a clear distinction between these species. The species or hybrid status of all accessions was determined. Principal coordinate analysis (PCA) based on pairwise genetic distance also identified four main genetic clusters corresponding to the four species (Figure 2).

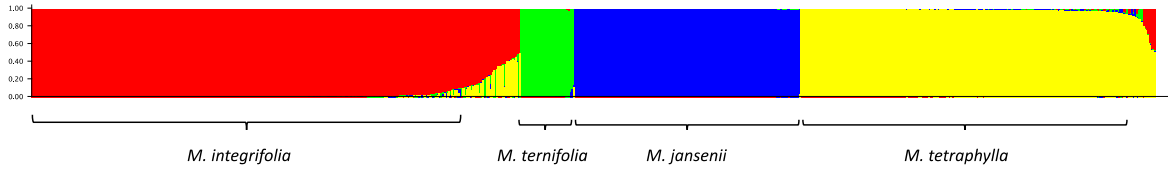

Figure 1. Barplot showing assignment of macadamia accessions to the four major genetic clusters identified in STRUCTURE. These correspond to the four *Macadamia* species. Y-axis is the proportion of membership to each species, and interspecific hybrids with mixed ancestry. Vertical lines represent individuals.

Estimates of genetic diversity including heterozygosity and allelic diversity were highest in *M. tetraphylla* (He = 0.76, Na = 15.73), followed by *M. integrifolia* (He = 0.73, Na = 13.82) and *M. ternifolia* (He = 0.72, Na = 10.73). *M. jansenii* had the lowest genetic diversity overall (He = 0.47, Na = 2.23).

Table 2. Genetic diversity in macadamia species and cultivars. N, number of accessions; Na, mean number of alleles; Ne, number of effective alleles, Ho, observed heterozygosity; He, expected heterozygosity

	N	**Na**	**Ne**	**Ho**	**He**
Macadamia species					
integrifolia	343	13.82	4.20	0.64	0.73
ternifolia	42	10.73	4.86	0.59	0.72
jansenii	179	3.64	2.23	0.43	0.47
tetraphylla	272	15.73	6.33	0.64	0.76
Cultivars					
Australian	35	8.00	4.27	0.75	0.75
Hawaiian	33	6.27	2.91	0.60	0.59
Other	11	6.00	3.28	0.64	0.65

Interspecific Hybrids

Overall, a small number of interspecific hybrids (68/1067) were identified, representing 6.6% of the accessions genotyped. Of these, 16 were cultivars and many others were planted trees. This suggests that natural hybridization in macadamia may be limited. Interspecific hybrids included *M. integrifolia-M. tetraphylla* (61), *M. integrifolia-M. ternifolia* (6) and one *M. integrifolia-M. ternifolia-M. tetrapylla* hybrid, cultivar HAES 791.

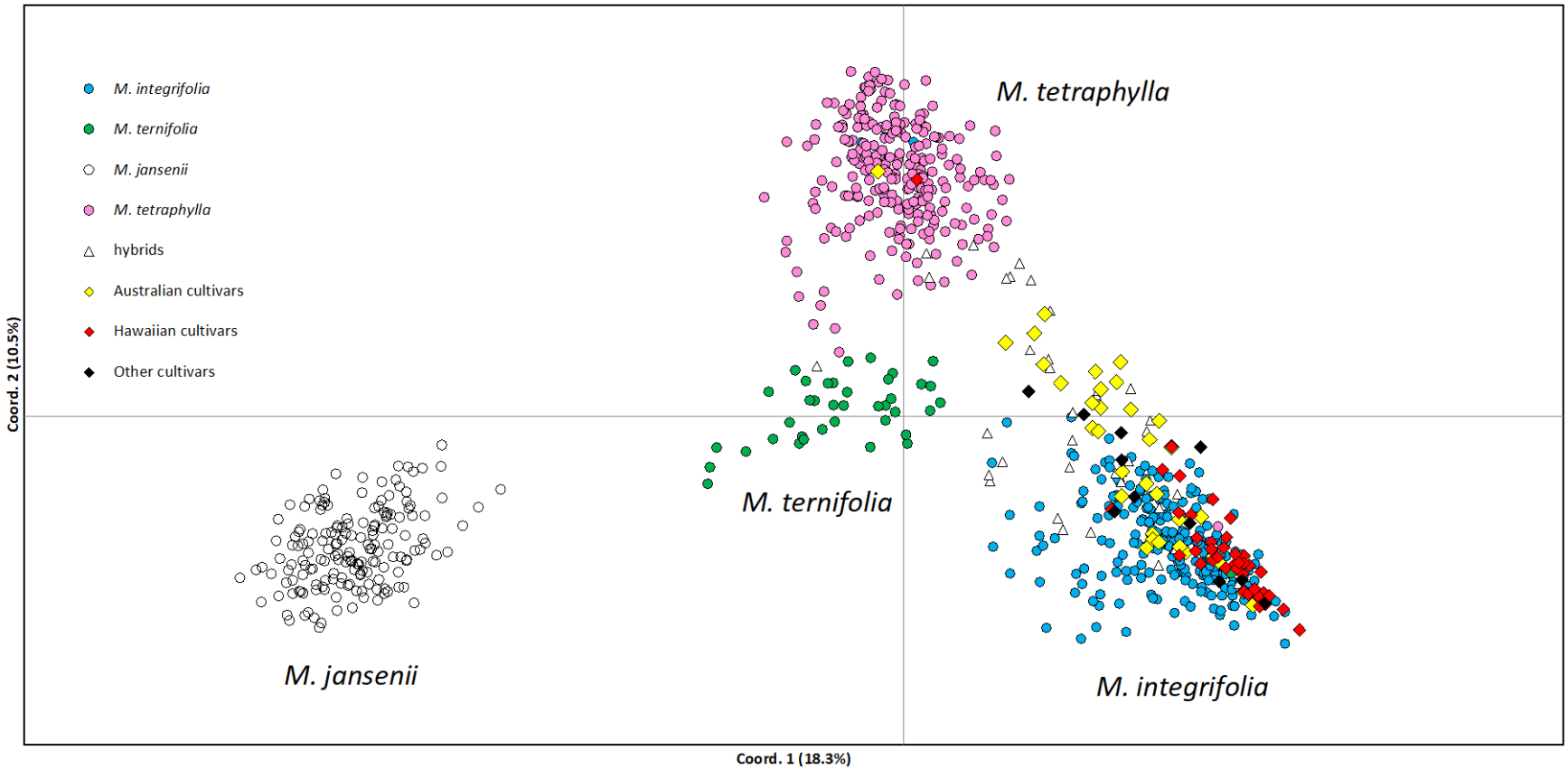

Figure 2. Principal coordinate analysis plot based on pairwise genetic distance among all macadamia accessions.

Macadamia integrifolia

Subsets of pure *M. integrifolia* (>90%, 343) and trees from presumed wild populations (160) were analysed to determine genetic structure. Five main genetic clusters, generally corresponding to geographic regions and chloroplast clades (C1-C5, Nock et al. 2019) were identified:

Northern group

1) Mt Bauple
2) Gympie region (Mooloo, Mary Creek, Amamoor)
3) North Brisbane (Villeneuve to Samford)
4) Brisbane (Holland Park, Sankey's Scrub to Mt Cotton)
5) South Brisbane to Gold Coast hinterland (Mt Cotton to Beechmont)

There was evidence of admixture between geographic regions that is the result of natural or human-mediated dispersal. For example, of the six sites surveyed at Mt Bauple, the two closest to the Highway included trees from the Brisbane and southern *M. integrifolia* groups suggesting that these sites included planted trees and their descendants. Admixture was greatest between regions in the southern group (Int 3-5, Figure 3).

Some sites showed little evidence of admixture and may represent original wild populations. These sites and individual trees assigning >80% to each of the genetic clusters could be considered when prioritizing populations for conservation, and selecting trees for inclusion in *ex situ* collections.

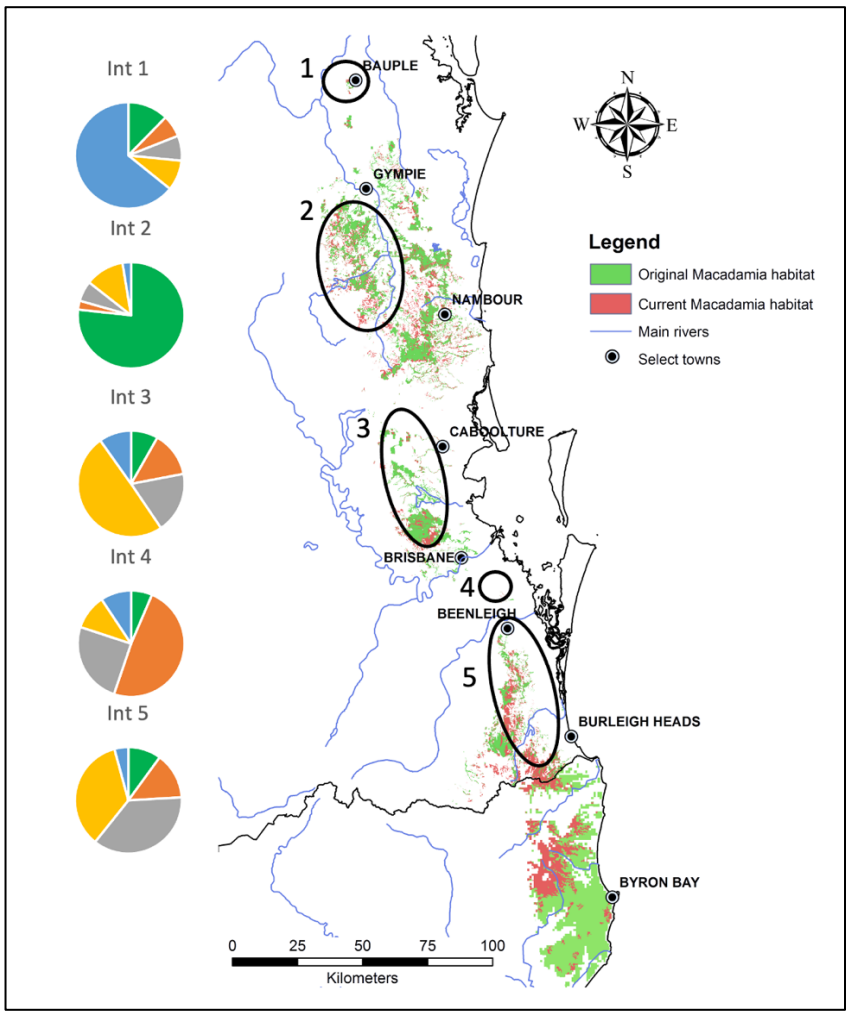

Figure 3. Geographic distribution of genetic diversity in *Macadamia integrifolia*. Approximate geographic groups (Int 1-5) are circled on the map. Pie charts show the proportional assignment of Int 1-5 to each of five genetic clusters identified in *M. integrifolia*.

Walter Hill tree

Planted in 1858, the Walter Hill (WH) tree at the Brisbane Botanical Gardens is one of the oldest recorded planted trees. Recent research based on chloroplast genome resequencing found that the WH tree was part of the southern group of *M. integrifolia* and formed a distinct clade only with other planted trees, including one at the University of California Berkeley and another in a Brisbane backyard (Nock et al. 2019). It was hoped that by using nuclear SSR markers (maternal and paternal) and including additional trees from the southern clade (Brisbane to the Gold Coast hinterland) that the wild origins of the WH tree could be identified.

Finding the wild origins of the WH tree is difficult because it is likely that seed from this tree and its source population has been planted widely. For example, the WH tree has a first-degree relationship with the tree at University of California, Berkeley USA planted in 1879, suggesting that the WH tree may be the seed parent. The WH tree is also closely related to a number of trees at Dulong and Montville. The Montville site is recorded as an old orchard planted from wild seed sourced in 1926, while the nearby Dulong site is recorded as wild *M. integrifolia*.

We found that among the presumed wild trees sampled, the WH tree was most closely related to trees from Sankey's Scrub, Holland Park, Brookside and Mt Cotton. At this stage, it seems probable that the WH tree was sourced from the Brisbane region. Sankey's Scrub, Holland Park and a Mt Cotton site (1023) may be remnants of what may have been a larger area of macadamia habitat prior to the development of Brisbane.

Walter Hill tree, Brisbane Botanical Gardens planted 1858

Macadamia tetraphylla

In total, 314 *M. tetraphylla* trees and 61 *M. tetraphylla-M. integrifolia* hybrids were genotyped. Overall genetic diversity was highest in *M. tetraphylla* compared to the other species (Table 1). Population structure was detected but, in comparison to *M. integrifolia,* there was greater admixture among distant sites. This is consistent with previous genetic research reporting a weak correlation between genetic and geographic distance in *M. tetraphylla* (O'Connor et al. 2015; Mai et al., 2020). Some sites showed little evidence of admixture and may represent remnant wild populations. These sites, and individual trees assigning >80% to each of the genetic clusters, could be considered when prioritizing populations for conservation, and selecting trees for inclusion in *ex situ* collections.

Five main genetic clusters were identified. These are represented by trees from the following sites:

1) Duroby
2) Mt Warning, Uki, Lismore, Nimbin, Guanaba
3) Lennox, Tyalgum
4) Victoria Park (MT10), Mt Glorious, Numulgi
5) Meerschaum Vale, Coolgardie, Victoria Park

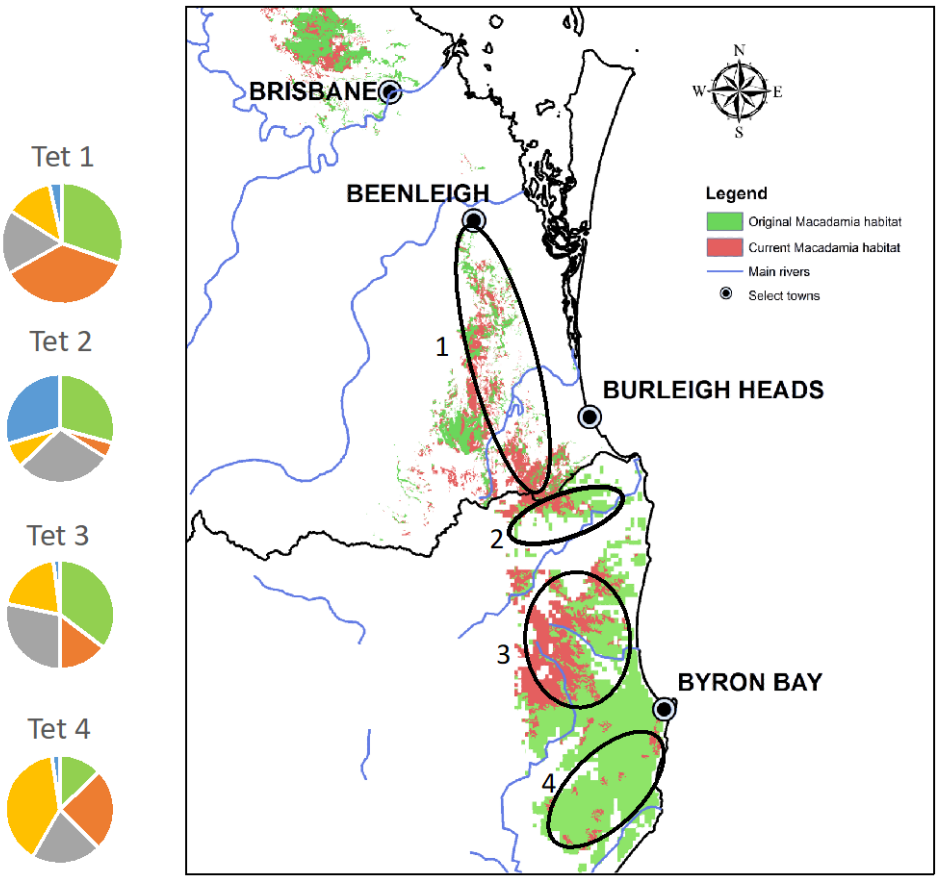

Figure 4. Geographic distribution of genetic diversity in *Macadamia tetraphylla*. Approximate geographic groups (Tet 1-5) are circled on the map. Pie charts show the proportional assignment of the geographic groups to each of five genetic clusters identified in *M. tetraphylla*.

Macadamia ternifolia

M. ternifolia is a small tree species of interest for future breeding or as a rootstock. The Hawaiian cultivar HAES 791 is cultivated in South Africa. 791 is a tri-species hybrid and the *M. ternifolia* content of HAES 791 was first identified by Cameron Peace (Peace et al. 2002).

In total, 40 *M. ternifolia* and 6 *M. ternifolia-M. integrifolia* hybrid trees were sampled and genotyped from the NMGC *ex situ* collection and populations identified through MC14000 *Still Wild about Macadamia*, the Wild Macadamia Hunt and Macadamia Conservation Committee survey activities. Genotyping of larger sample of *M. ternifolia* (40) and *M. jansenii* (179) trees in this project enabled genetic discrimination between these species that was not possible with fewer than 10 individuals (Mai et al. 2020).

M. ternifolia formed a distinct species cluster in STRUCTURE (Fig 1) and PCA (Fig 2) analyses. Allelic diversity and heterozygosity were lower than in *M. tetraphylla and M. integrifolia* but significantly higher than *M. jansenii* (Table 2). Too few trees were sampled across the distribution to effectively examine population structure in *M. ternifolia.* This species is underrepresented in *ex situ* collections and further work is needed to understand the distribution of genetic diversity in this species.

Clockwise from top left, Russell Bennet with *M. ternifolia* nuts. Ben and Renee from Australian Plant Specialists collecting *M. ternifolia* scion for the macadamia arboretum. Paul O'Hare, Ian McConachie and Janelle Vaughan with *M. ternifolia* tree.

Macadamia jansenii

Collaboration with Professor Alison Shapcott and Honours student Glenn Hayward, University of Sunshine Coast provided the opportunity to genotype leaf samples from all known *M. jansenii* trees (182). These genetic data were used to support Glen Hayward's Honours research that was published in Australian Journal of Botany https://www.publish.csiro.au/BT/BT20160. Hayward et al. (2021) found 89% of the allelic variation in *M. jansenii* was captured in the existing *ex situ* collections. When included in an analysis with all other accessions from this study *M. jansenii* formed a distinct species cluster (Fig 1 and 2). Genetic diversity was in *M. jansenii* was significantly lower in comparison to other *Macadamia* species (Table 2).

Cultivated Varieties

Cultivars were identified as species or hybrids based on STRUCTURE and PCA analyses. *Macadamia integrifolia* is the main species used in cultivation. There are close relationships (at or above the level expected between full-sibs and parent-offspring) among most Hawaiian cultivars. Australian cultivars are derived from more diverse germplasm and many are hybrids of *M. integrifolia* and *M. tetraphylla* (Fig. 2, Fig. 5).

Among the cultivated varieties, genetic diversity was highest in the Australian cultivars. The level of expected heterozygosity (He = 0.75) in the Australian cultivars was similar to that in *M. integrifolia* and *M. tetraphylla,* the species from which these cultivars were derived. Allelic diversity was also highest in the Australian cultivars, in comparison to those developed in Hawaii (He = 0.59) and other countries (Table 2).

Figure 5. Macadamia cultivars and species status based on SSR genotype data. *M. integrifolia* and *M. tetraphylla* cultivars have >90% content from one species. Hybrid* are first-generation hybrids of *M. integrifolia* x *M. tetraphylla.* Other hybrids have <50% *M. tetraphylla* content. On right, bar plot showing proportion of *M. integrifolia* (blue), *M. tetraphylla* (green) and *M. ternifolia* (orange) in each of the cultivars.

Origins of domestication

Hawaiian cultivars

Genetic evidence from this project indicates that the most likely wild origin of the Hawaiian cultivars is Mt Bauple. Prior evidence from maternally inherited chloroplast data suggested that they originated from Mooloo in the Gympie region and Mt Bauple (Nock et al. 2019). SSR and whole genome resequencing data collected as part of this project, and collaborative research with international colleagues, provides more detailed information from the nuclear genome (maternal and paternal inheritance) and presents a clearer picture of the domestication history of macadamia. Interestingly, there was evidence of maternal gene flow (seed dispersal) from the Gympie region to Mt Bauple. Given that small rodents, gravity and water are the proposed natural mechanisms for seed dispersal it is most likely that humans were responsible for the movement of seed. So, although the original maternal (chloroplast) lineage of many Hawaiian cultivars traces to Mooloo, the seed collected in Australia and sent to Hawaii in the late 19th century was most likely collected at Mt. Bauple.

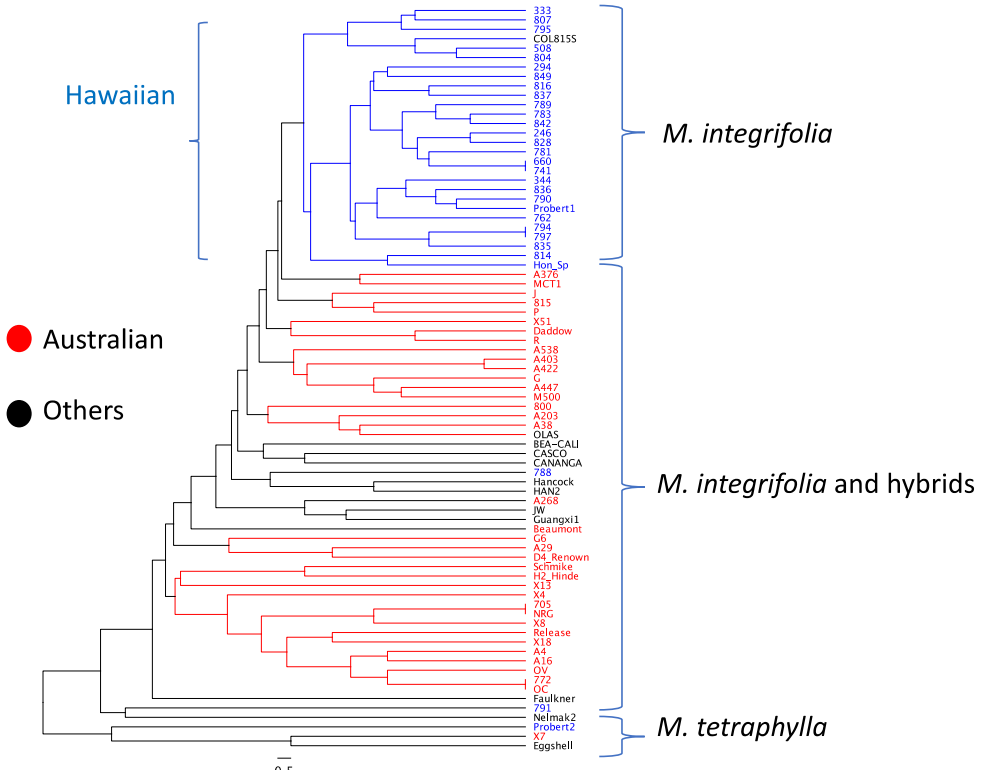

Figure 6. UPGMA dendrogram based on SSR genotype data. Hawaiian cultivars in blue, Australian cultivars in red.

Australian cultivars

Of the 35 Australian cultivars included in this study, 16 were identified as *M. integrifolia* x *M. tetraphylla* hybrids including four first generation interspecific hybrids: Norm Greber's X4 and X7, Renown, Beaumont and HVP A268 (Figure below). Eggshell is an *M. tetraphylla* cultivar.

Renown

Renown, or D4, is a parent of A4 and A16 and may be the main source of *M. tetraphylla* germplasm in the Australian cultivars, particularly those developed by Hidden Valley Plantations (HVP). SSR data confirm that Renown, or D4, is a first-degree relative (parent-offspring or full-sib) of cultivars A4, A16, A203, A29 and Norm Greber's X8. It is also a first-degree relative of a tree from Wallaville, near Bundaberg estimated to be >100 years old. The Gympie region is the most likely origin of the *M. integrifolia* parent of Renown. It is a second-degree relative (grandparent-grandchild or half-sib) of wild trees from Mooloo and Mary's Creek. Of the sites sampled for this study, Crystal Creek in NSW is the most likely origin of the *M. tetraphylla* parent of Renown.

Own Choice

Own Choice (OC) is reportedly the *M. integrifolia* parent of A4 and A16 from Hidden Valley Plantations (Hardner et al. 2009). The SSR data support this, and confirm that OC has the same genotype as HAES 772. OC is a first-degree relative of A4, A16, Own Venture and is also related to A538, X18, X4 and A447. OC is closely related to wild trees from Mooloo and Amamoor suggesting that it originated in the Gympie region. This is concordant with historical records suggesting that some of Norm Greber's selections originate from seed collected in the Amamoor Creek Valley (Trochoulias et al. 1989, Vithanage and Winks 1992, Hardner et al. 2009).

Beaumont

Beaumont (HAES 695, NSW-44) is a first-generation *M. integrifolia* x *M. tetraphylla* hybrid. It is most closely related to wild *M. tetraphylla* trees at Tyalgum and Crystal Creek. This cultivar was rejected in Hawaii but later taken to California where it was named Beaumont (Hardner et al. 2009). It is a likely parent of the South

American cultivars Casco and Tambor. Beaumont was found at presumed wild sites at Nimbin and Dulong, suggesting that these may not represent wild populations.

Daddow

Daddow is one of few cultivars from the southern distribution of *M. integrifolia,* as first suggested by the early genetic research of Cameron Peace (Peace et al. 2002). Genotype data confirm that Daddow is a parent of two recent cultivars form the Australian macadamia breeding program, R and G (Alam et al. 2018).

Outcomes

This project is aligned with priorities outlined in the Macadamia Strategic Investment Plan 2017-2021. In particular, it contributes to Outcome 2, Improved production systems covering plant breeding, intensive orchards and novel technologies and the Key Performance indicator, Progress has been made with the conservation of wild macadamia stocks through levy support for the Macadamia Conservation Trust.

The primary outcome was identification of the distribution and structure of genetic diversity in wild, planted and domesticated germplasm. This information will help in the prioritisation of populations and trees for conservation and inclusion in future *ex situ* collections. Genotyping of >1000 trees from sites of wild, uncertain and planted origin provided evidence of remnant population structure, and also extensive translocations throughout the original habitat of macadamia. This is not surprising given the historical records of early orchards and planted trees (McConachie 2012), and that macadamia nuts were likely an important and tradable food for Aboriginal people (Hardner et al. 2009, Kerkhove 2013).

Macadamia integrifolia and *M. integrifolia-M. tetraphylla* hybrids provide the basis of the global macadamia industry. Most Hawaiian cultivars are *M. integrifolia* with the exception of HAES 741, a tri-species hybrid with *M. ternifolia* content and the *M. tetraphylla* cultivar Probert 2. In contrast, the Australian cultivars including those developed by Norm Greber and Hidden Valley Plantations have a more diverse ancestry and include many hybrids. It is interesting that many Australian cultivars (P, J, MCT1, A268, A203, A4, A16) and cultivars of importance in South Africa (Beaumont, 791), China (JW, Quangxi 1) and Guatemala (Casco and Cananga) are interspecific hybrids.

Wild populations of *M. tetraphylla*, *M. ternifolia* and the southern distribution of *M. integrifolia* contain novel genetic diversity that is yet to be incorporated into breeding programs. The conservation of key natural populations of all *Macadamia* species will be important for future improvement of this native Australian nut crop. Most cultivars are derived from the northern *M. integrifolia* group at Mt Bauple and the Gympie region. The southern *M. integrifolia* group and *M. tetraphylla* represent a mostly unexploited source of genetic and morphological diversity that could drive future productivity.

This project provides new information on genetic diversity and structure of macadamia that can be used to prioritise wild populations and individual trees for conservation. It is expected this will be further refined as more whole genome resequencing data becomes available.

Monitoring and evaluation

The project developed a program logic and a monitoring and evaluation plan as part of its project monitoring and evaluation activities. The below Key Evaluation Questions were developed at commencement and reviewed at the end of the project:

Key Evaluation Questions (KEQ)	End of project review
1. To what extent has the project achieved its expected outcomes?	The project has identified the distribution and structure of genetic diversity in wild, planted and domesticated germplasm.
2. How relevant was the project to the needs of intended beneficiaries?	The new knowledge generated in the project will help in the prioritisation of populations and trees for conservation and inclusion in future *ex situ* collections. It is expected that the results of this project will be beneficial in the short-term for macadamia breeders, nurseries and growers and assist in the conservation of wild populations in the long-term for the benefit of the macadamia industry globally (1-100 years).
3. Have regular project updates been provided through linkage with the industry communication project?	Throughout the project, the team engaged with external stakeholders including growers, landowners, MC19000 breeding program, AS17000 National Tree Genomics Program, universities, researchers, Australian Macadamia Society and Macadamia Conservation Trust constructively to achieve the project outcomes. The massive effort to sample >1000 trees was only possible through engagement with the Macadamia Conservation Committee, Healthy Land and Water Qld, citizen science project, the Wild Macadamia Hunt, macadamia breeders and growers, and private property owners.

Overall, the project achieved its aim of identifying the distribution of genetic diversity in wild and domesticated macadamia. A large collection of macadamia accessions including >1000 wild and planted trees representing all macadamia species and important cultivars was sampled from across the distribution of macadamia and genotyped through this project. Genotype data was used to determine the species or hybrid status of all accessions, and detect population structure in *M. integrifolia* and *M. tetraphylla.*

Identification of the major genetic clusters within macadamia is important for the prioritisation of wild populations for conservation and the selection of trees for inclusion in *ex situ* germplasm plantings and breeding programs. It is expected that the results of this project will be beneficial in the short-term for macadamia breeders, nurseries and growers and assist in the conservation of wild populations in the long-term for the benefit of the macadamia industry globally (1-100 years). While it is probable that the genetic markers implemented in this project will be superseded in future, SSRs are currently the only multi-allelic markers available and remain widely used in Australia and internationally in agriculture and human forensics. As such, these macadamia-specific markers are likely to have continued use in the medium to long-term (10+ years) for DNA authentication and paternity analysis to confirm breeding crosses and to trace gene flow in orchards.

The target audience for adoption of the outcomes of this work are the Australian Macadamia industry and the Macadamia Conservation Trust. Collaboration with the project teams from other Hort Innovation funded macadamia programs, and engagement with the Macadamia Conservation Trust and the macadamia industry ensured that the information generated was relevant and useful for future breeding and conservation. The

project links to other Hort Innovation funded projects including MC15008 *Unravelling the genetics of macadamia: integration of linkage and genome maps*, MC14000 *Still Wild about Macadamia,* MC14000 *Macadamia Second Generation Breeding and Conservation* and AS17000 *National Tree Genomics Program*.

Throughout the project, the team engaged with external stakeholders including growers, landowners, MC19000 breeding program, AS17000 National Tree Genomics Program, universities, researchers, Australian Macadamia Society and Macadamia Conservation Trust constructively to achieve the project outcomes. The massive effort to sample >1000 trees was only possible through engagement with the Macadamia Conservation Committee, Healthy Land and Water Qld, citizen science project, the Wild Macadamia Hunt, macadamia breeders and growers, and private property owners.

Despite the restrictions imposed by the COVID-19 pandemic, regular contact was maintained throughout the project with collaborators Bruce Topp and Craig Hardner, University of Queensland. Preliminary results were presented at the international TropAg conference in Brisbane, November 2019 and meetings of the Macadamia Industry Variety Improvement Committee (MIVIC). Articles were published in the Australian Macadamia Society news bulletin (Appendix 1 and 2).

COVID restrictions to cross border travel limited the field trips that were planned as part of this project. 'Remote' collection was crucial to project success. The project team worked with Healthy Land and Water, and the Wild Macadamia Hunt https://hlw.org.au/project/the-wild-macadamia-hunt/ to optimise remote sample and data collection for citizen scientists. Sample collection kits and advice helped to ensure that DNA was preserved and suitable for genotyping. These methods were also provided to the Macadamia Conservation Trust for collection under scientific license.

The SSR markers employed in this project could be of benefit for breeding programs as a simple, and relatively cheap method to confirm paternity of seedlings, variety identification. Subsets of SSR markers were identified for paternity analysis (11, Langdon et al. 2019), and variety identification (9) and these have been shared with the Australian breeding program. Methods to reduce the per sample cost of genotyping included optimised methods for sample collection and preservation, DNA extraction and genotyping were developed and adopted throughout the project. These efficiencies made it possible to maximise the number of wild populations and accessions genotyped.

Recommendations

1) Macadamia is a native Australian tree nut crop. Further research by Hort Innovation into key wild populations and trees that represent distinct genetic clusters in each species is needed to identify, understand and conserve valuable genetics with the potential to drive future industry productivity.
2) Additional surveys and further research on *M. ternifolia* are needed to identify population structure and natural hybrid zones. While this was beyond the scope of this project, many new populations were identified and the genetic distinctiveness of this species was confirmed.
3) Trees representing most of the distinct genetic clusters identified in *M. integrifolia* and *M. tetraphylla* are represented in the National Macadamia Germplasm *ex situ* collection. However, long-term conservation of key wild populations is also required to ensure that the extent of natural genetic and morphological variation in macadamia is available for the global macadamia industry. These efforts will also assist in safeguarding the critically endangered lowland subtropical rainforests ecosystems that are home to macadamia.

Refereed scientific publications

Hayward, G., Nock, C.J. Shimizu, Y., Shapcott, A. 2021. Demographic, population genetics, population viability analysis, and habitat model assist in the assessment and future planning of a reintroduction of the endangered rainforest species, *Macadamia jansenii* (Proteaceae) post fire. *Australian Journal of Botany*. 10(10): 3497-3504 https://www.publish.csiro.au/BT/BT20160

Nock, C.J., Baten, A., Mauleon, R., Langdon, K.S., Topp, B., Hardner, C.M., Furtado, A., Henry, R.J. and King, G.J., 2020. Chromosome-scale assembly and annotation of the macadamia genome (*Macadamia integrifolia* HAES 741). *G3: Genes, Genomes, Genetics* 10: 3497-3504 https://www.g3journal.org/content/10/10/3497

Topp, B.L., Nock, C.J., Hardner, C.M., Alam, M.M. and O'Connor, K.M., 2020. Macadamia (*Macadamia spp.*) Breeding. In: Al-Khayri J., Jain S., Johnson D. (eds) Advances in Plant Breeding Strategies: Nuts and Beverage Crops. Springer, Cham, https://link.springer.com/chapter/10.1007/978-3-030-23112-5_7

Sharma, P., Murigneux, V., Haimovitz, J., Nock, C., Tian, W., Kharabian-Masouleh, A., Topp, B., Alam, M., Furtado, A. and Henry, R., in press. The genome of the endangered *Macadamia jansenii* displays little diversity but represents an important genetic resource for plant breeding. *Plant Direct,* Accepted 10/11/2021.

Conference Presentation

Nock, C.J., Langdon, K.S., Baten, A., Mauleon, R., Hardner, C.M., O'Connor, K., Topp, B. and King, G.J. (2019) Genomics and the macadamia orchard of the future. TropAg, Brisbane November 11-13th.

References

Alam, M.M., Hardner, C.M., Nock, C.J., O'Connor, K.M. and Topp, B.L., 2019. Historical and molecular evidence of genetic identity of HAES 741 and HAES 660 macadamia cultivars. *Hort Science* 54(4):616-620

Brozynska, M., Furtado, A. and Henry, R.J., 2016. Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant biotechnology journal*, *14*(4), pp.1070-1085.

Hardner, C.M., Peace, C., Lowe, A.J., Neal, J., Pisanu, P., Powell, M., Schmidt, A., Spain, C. and Williams, K., 2009. 1 Genetic Resources and Domestication of Macadamia. *Horticultural reviews*, *35*(1).

Hayward, G., Nock, C.J. Shimizu, Y. and Shapcott, A., 2021. A comprehensive approach to assessing the future persistence of the endangered rainforest tree, Macadamia jansenii (Proteaceae) and the impact of fire. *Australian Journal of Botany* 69(5): 285-300

Kamper, W., Trueman, S.J., Ogbourne, S.M. and Wallace, H.M., 2021. Pollination services in a macadamia cultivar depend on across-orchard transport of cross pollen. *Journal of Applied Ecology*, *58*(11), pp.2529-2539.

Kerkhove, R., 2013. Aboriginal trade in fish and seafoods to settlers in nineteenth-century South-East Queensland: A vibrant industry? *Queensland Review*, *20*(2), pp.144-156.

Langdon, K.S., King, G.J. and Nock, C.J., 2019. DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. *Tree Genetics and Genomes* 15:29.

Mast, A. R., Willis, C. L., Jones, E. H., Downs, K. M., and Weston, P. H., 2008. A smaller Macadamia from a more vagile tribe: inference of phylogenetic relationships, divergence times, and diaspore evolution in Macadamia and relatives (tribe Macadamieae, Proteaceae). *American Journal of Botany* 95: 843–870.

Mai, T., Alam, M., Hardner, C., Henry, R. and Topp, B., 2020. Genetic Structure of Wild Germplasm of Macadamia: Species Assignment, Diversity and Phylogeographic Relationships. *Plants*, *9*(6), p.714.

McConachie, I., 2012. The Macadamia story – an overview. Australian Macadamia Society, Lismore

Nock, C.J., Elphinstone, M.S., Ablett, G., Kawamata, A., Hancock, W., Hardner, C.M. and King, G.J., 2014. Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild Macadamia (Proteaceae). *Applications in Plant Sciences* 2(4): 1300089.

Nock, C.J., Hardner, C.M., Montenegro, J.D., Ahmad Termizi, A.A., Hayashi, S., Playford, J., Edwards, D., Batley, J., 2019. Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. *Frontiers in Plant Science* 10, 334.

O'Connor, K., Powell, M., Nock, C., and Shapcott, A. (2015). Crop to wild gene flow and genetic diversity in a vulnerable Macadamia (Proteaceae) species in New South Wales, Australia. *Biological Conservation* 191, 504–511.

Peace, C. P., 2002. Genetic characterisation of Macadamia with DNA Markers. PhD. Thesis. University of Queensland.

Trochoulias et al. 1989; An historical perspective on variety evaluation in Australia from interviews with Norm Greber, patron of the Society. *Australian Macadamia Society News Bulletin* 16, 17–19.

Vithanage, V., and C.W. Winks. 1992. Isozymes as genetic markers for macadamia. *Scientia Horticulturae* 49, 103–115.

Intellectual property, commercialisation and confidentiality

This report is intended for dissemination. Appendixes 3, 4 and 5 are to be maintained confidential.

No project IP, project outputs, commercialisation or confidentiality issues to report.

Acknowledgements

The MC18004 project team included Catherine Nock, Ramil Mauleon and Asuka Kawamata from Southern Cross University. We thank project collaborators Bruce Topp and Craig Hardner from University of Queensland for their contributions, and Hort Innovation project manager Vino Rajandran for his support and advice throughout.

This project, and the sampling of >1000 accessions was possible because of the major contributions of a large group of organisations and individuals. We thank the Macadamia Conservation Trust, Australian Macadamia Society, Healthy Land and Water Qld, University of Queensland, University of Sunshine Coast, NSW Department of Primary Industries, Queensland Department of Agriculture and Fisheries, University of Adelaide, USDA, Hidden Valley Plantations and the Australian macadamia breeding program. Many thanks to Ian McConachie, Brice Kaddatz, Paul O'Hare, Denise Bond, Jolyon Burnett, Liz Gould, Vanessa Durand, Alison Shapcott, Glenn Hayward, Michael Powell, David Bell, Michelle Wirthensohn, Katie O'Connor, Mobashwer Alam, Dougal Russell, Ken Dorey, Andy Burnside, Russell Bennet, Craig Maddox, Scott Ellem and Tim Roberts. Thank you also to the many private landholders for sharing their knowledge and and permission to sample trees, and to the citizen scientists for their participation and support of Healthy Land and Water's Wild Macadamia Hunt.

Appendices

Appendix 1. Nock, C. Genetic variation in macadamia cultivars. AMS News Bulletin, Summer 2020

Appendix 2. Hardner C, Peace C, Nock, C and McConachie I. Short history of macadamia cultivars. AMS News Bulletin, Summer 2020