



Wild Origins of Macadamia Domestication Identified Through Intraspecific Chloroplast Genome Sequencing

Catherine J. Nock^{1*†}, Craig M. Hardner^{2†}, Juan D. Montenegro³, Ainnatul A. Ahmad Termizi², Satomi Hayashi⁴, Julia Playford⁵, David Edwards⁶ and Jacqueline Batley⁶

¹ Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia, ² Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia, ³ Australian Genome Research Facility, St Lucia, QLD, Australia, ⁴ Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, QLD, Australia, ⁵ Queensland Department of Environment and Science, Brisbane, QLD, Australia, ⁶ School of Biological Sciences and Institute of Agriculture, The University of Western Australia, Crawley, WA, Australia

OPEN ACCESS

Edited by:

Kathleen Pryer,
Duke University, United States

Reviewed by:

Eric Wade Linton,
Central Michigan University,
United States
Alessandro Alves-Pereira,
Campinas State University, Brazil

*Correspondence:

Catherine J. Nock
cathy.nock@scu.edu.au

† These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 19 September 2018

Accepted: 04 March 2019

Published: 21 March 2019

Citation:

Nock CJ, Hardner CM,
Montenegro JD, Ahmad Termizi AA,
Hayashi S, Playford J, Edwards D and
Batley J (2019) Wild Origins
of Macadamia Domestication
Identified Through Intraspecific
Chloroplast Genome Sequencing.
Front. Plant Sci. 10:334.
doi: 10.3389/fpls.2019.00334

Identifying the geographic origins of crops is important for the conservation and utilization of novel genetic variation. Even so, the origins of many food crops remain elusive. The tree nut crop macadamia has a remarkable domestication history, from subtropical rain forests in Australia through Hawaii to global cultivation all within the last century. The industry is based primarily on *Macadamia integrifolia* and *M. integrifolia*–*M. tetraphylla* hybrid cultivars with Hawaiian cultivars the main contributors to world production. Sequence data from the chloroplast genome assembled using a genome skimming strategy was used to determine population structure among remnant populations of the main progenitor species, *M. integrifolia*. Phylogenetic analysis of a 506 bp chloroplast SNP alignment from 64 wild and cultivated accessions identified phylogeographic structure and deep divergences between clades providing evidence for historical barriers to seed dispersal. High levels of variation were detected among wild accessions. Most Hawaiian cultivars, however, shared a single chlorotype that was also present at two wild sites at Mooloo and Mt Bauple from the northernmost distribution of the species in south-east Queensland. Our results provide evidence for a maternal genetic bottleneck during early macadamia domestication, and pinpoint the likely source of seed used to develop the Hawaiian cultivars. The extensive variability and structuring of *M. integrifolia* chloroplast genomic variation detected in this study suggests much unexploited genetic diversity is available for improvement of this recently domesticated crop.

Keywords: *Macadamia integrifolia*, Proteaceae, chloroplast phylogenomics, crop domestication, phylogeography, bottleneck

INTRODUCTION

Understanding the relationships between domesticated and wild germplasm is important to guide introduction of novel genetic diversity into selective breeding populations, and to prioritize conservation of novel wild germplasm that may be useful in the future (Brozynska et al., 2016; Chen et al., 2017; Luo et al., 2017; Zhang et al., 2017). Most major crops are derived from northern

hemisphere Monocotyledon (monocot) and core Eudicotyledon (eudicot) species that were first domesticated 1000s of years ago (Miller and Gross, 2011). For these crops, there have generally been few domestication events and a small portion of the available genetic diversity in the wild progenitor species was selected (Wright et al., 2005; Doebley et al., 2006; Haudry et al., 2007; Meyer and Purugganan, 2013). A long history of selection, dispersal, hybridization and introgression, can lead to divergence between domesticated and wild source germplasm often obscuring the geographic origins of domestication (Burger et al., 2008; Fuller et al., 2011; Meyer et al., 2012). In addition, depending upon the intensity of anthropogenic activity, the original populations may be disturbed or lost (Fuller et al., 2011). In contrast, for more recently domesticated crops, there is the potential to identify specific source populations, although pinpointing the geographic origins of domestication requires a detailed knowledge of the population structure of the progenitor species (Schmutz et al., 2014).

Macadamia is unique in comparison to other horticultural tree crops. *Macadamia* (F. Muell, $2n = 28$) is a subtropical rain forest genus in the Proteaceae, an early-diverging eudicot family that had diversified in Australia by the Late Cretaceous (Mast et al., 2008; Sauquet et al., 2009; Nock et al., 2014; Carpenter et al., 2015). The four species in the genus are endemic to the lowland subtropical rain forest of eastern Australia and have a discontinuous distribution from south-east Queensland to north-east New South Wales (Powell et al., 2010, 2014). *Macadamia* is one of few international food crops derived from either the basal eudicots or the flora of Australia. Two species, *Macadamia integrifolia* and *M. tetraphylla*, produce an edible high-value oil rich kernel. Although it was likely a component of the diet of the indigenous peoples of Australia, to our knowledge, there is no recorded evidence of cultivation prior to European occupation of the natural habitat of the genus in the mid 19th century (Gross, 1995; Costello et al., 2009; Hardner et al., 2009). The first European contact with the genus was reportedly in 1848 (Smith, 1956) and the first cultivated macadamia may be a tree planted in 1858 by Walter Hill in the Brisbane Botanical Gardens that is still alive today. Early botanists exported macadamia seed in the mid to late 19th century while the first orchards in Australia were established from the late 19th century most likely with germplasm from proximally located native forest (McConachie, 1980; Hardner et al., 2009).

The expansion of macadamia as a commercial crop initially occurred in Hawaii from the 1920s (Wagner-Wright, 1995; Hardner et al., 2009; Hardner, 2016). The favored species for commercial production, *M. integrifolia*, was initially introduced to Hawaii in two separate events in the late 19th century. The first introduction was by W. H. Purvis sometime between 1881 and 1885, with trees planted near Kukuihaele on the Big Island. Subsequently, R. A. Jordan introduced macadamia into Hawaii in 1892 with trees from this second introduction grown in Honolulu on Oahu (Hardner, 2016). With recognition of the eating quality of the kernel, commercial seedling orchards were established throughout the Hawaiian Islands from the 1920s with seedlings trees reportedly derived directly from the 19th century introductions (Shigeura and Ooka, 1984;

Hardner, 2016). Following the development in Hawaii of reliable grafting techniques in the mid 1930s, seedling orchards were surveyed to identify elite trees that were subsequently clonally propagated. The performance of selected cultivars was evaluated prior to commercial release to the industry for the establishment of new plantations (Hardner et al., 2009).

The macadamia industry has undergone rapid global expansion in the last 50 years. Australia, South Africa, Kenya, and United States are currently the largest producers and the crop is also cultivated in China, southeast Asia, South America, Malawi, and New Zealand. Future growth in global production is predicted following recent extensions in planting, particularly in China and South Africa¹. A few pure *M. tetraphylla* cultivars are grown commercially in South Africa (Peace et al., 2005). However, most industry cultivars are *M. integrifolia* or hybrids of *M. integrifolia* and *M. tetraphylla*. The *M. integrifolia* cultivars developed in Hawaii account for the majority of current world production and important founders of current breeding programs (Hardner, 2016). Knowledge of the extent and structure of genetic diversity is important for future genetic improvement, particularly in crops such as macadamia which is clonally propagated. *Macadamia* is adapted to subtropical rain forest habitat and recent genomic evidence points to an expansion of gene families involved in plant defense and pathogen recognition (Nock et al., 2016). A wide range of pests and diseases impact macadamia productivity and the identification of population structure and natural genetic variation is likely to be important in the development of resistant varieties.

The chloroplast is a plant organelle originating from an ancestral free-living cyanobacterium through endosymbiosis and performs a fundamental role in plant metabolism including photosynthesis (Gray and Doolittle, 1982; Timmis et al., 2004). The structure and gene content of chloroplast genomes are generally well-conserved among photosynthetic land plant species. They contain a large single copy (LSC) and small single copy (SSC) region separated by two inverted repeat (IR) sequences and range in size from 107 to 218 kb (Palmer, 1991). In contrast to the bi-parentally inherited nuclear genome, the chloroplast genome of most flowering plants is maternally inherited without recombination. Consequently, the chloroplast genome has been particularly useful for studying the maternal evolutionary history, or seedline, of angiosperms. Until relatively recently, intraspecific studies in particular were based on limited variation found in short PCR-amplifiable regions of the genome (Taberlet et al., 1991; Hamilton, 1999; Provan et al., 2001; Shaw et al., 2007). The development of next generation sequencing (NGS) technologies has led to a massive increase in the availability of shotgun sequence data for many plant species. This enables recovery of whole chloroplast genome sequences using a range of different techniques. These include (i) assembly of long-range PCR amplicons (Cronn et al., 2008; Whittall et al., 2010), (ii) 'genome skimming'- shallow sequencing of total DNA that provides deep sequencing of high-copy chloroplast DNA, followed by assembly to a reference genome (Nock et al., 2011; Straub et al., 2012; Bock et al.,

¹www.nutfruit.org/consumers/news/detail/inc-2017-2018-statistical-yearbook

2014; Dodsworth, 2015) and (iii) *de novo* assembly with deeper NGS read coverage where no reference sequence is available (McPherson et al., 2013; Izan et al., 2017). Variation in the chloroplast genome has provided important insights into the domestication origins of crops including apple and citrus (Nikiforova et al., 2013; Carbonell-Caballero et al., 2015; Daniell et al., 2016). Recently, the complete chloroplast genome *M. integrifolia* cultivar HAES 741 'Mauka' was sequenced (Nock et al., 2014) enabling comparative analysis of chloroplast variability assessed through a genome skimming strategy. In this study, intraspecific chloroplast sequence variation is used to investigate the population structure of remnant macadamia germplasm and applied to infer the origins of macadamia domestication. In contrast to other perennial tree crops species, this may be feasible due to the persistence of many wild populations and the short domestication history of macadamia.

MATERIALS AND METHODS

Plant Material and DNA Extraction

The National Macadamia Germplasm Conservation Program established *ex situ* plantings of clones of wild trees sampled in 1996 as cuttings from naturally occurring populations comprising most of the geographic range of the four species (Peace, 2002; Hardner et al., 2004). These *ex situ* plantings, located at Alstonville, Tiaro, and Burpengary in eastern Australia, were the source of most of the wild germplasm included in this study. The 64 accessions in total included (i) 37 samples from wild populations spanning the geographical distribution of *M. integrifolia* (Table 1A) (ii) cultivated germplasm including 26 *M. integrifolia* cultivars, selections, and cultivated trees (Table 1B) and (iii) a sample of *M. janseni* as an outgroup in phylogenetic analyses. The wild accessions were originally sampled from 26 sites that were clustered into localities. A map of the predicted remnant distribution of macadamia was produced following habitat mapping methods outlined in Powell et al. (2010) to display the geographic location of the original sites sampled (Figure 1).

With the exception of W30.WV, which was sampled from the original remnant tree, all wild accessions were sampled from the *ex situ* conservation plantings. The wild accession from Willowvale (W30.WV) is considered a maternal source of the Jordan introduction of *M. integrifolia* into Hawaii in the late 19th century from which most of the initial seedling orchards are thought to have been derived (Hardner, 2016). Putatively planted and hybrid accessions suggested through collection notes and earlier molecular analyses (Peace, 2002) were not included in this study.

The 26 samples of cultivated germplasm included 11 Hawaiian cultivars: nine that were originally propagated directly from the early seedling orchards and are referred to using the Hawaiian Agricultural Experimental Station selection numbers (H.246, H.294, H.333, H.344, H.425, H.508, H.660, H.741, H.791), a cultivar selected by the Honokaa Sugar Company from their original seedling orchard (Honokaa Special, H.HSp), and an open pollinated selection derived from this cultivar

(H.814). Other cultivated samples included three trees that are considered to be derived from the Purvis introduction (H.Purv, H.TPN, and H.TPS), a putative relative of the original Jordan introduction (H.Cwy), five samples from the Nutridge seedling orchard established in the 1920s near Honolulu (H.Nut03, H.Nut07, H.Nut12, H.Nut14, H.Nut15) from which five cultivars originated, and a seedling tree growing in the Waipio Valley believed to be from an old seedling orchard planted there in the 1930s (H.Wai) (Hardner, 2016). Three samples of old seedling trees from California were also included, a tree (C.UCB) planted in the 1879 the campus of the University of California, Berkeley (Storey, 1977), a sample from a scion of an old macadamia cultivar Faulkner (C.Fau) that was selected from a planting at Santa Paula, CA that had been propagated from seed introduced from Florida about 1900 (Schroeder, 1954), and a tree (C.Hei) growing on the Coronado peninsula, San Diego that was planted about 1890 (Trask, 1962). Two samples were obtained from cultivated trees in Brisbane Australia, including the Walter Hill Tree planted in 1858 (A.WH) and a tree growing in the backyard of the suburb of Yeronga (A.Yer) planted approximately 60–70 years ago. A *M. janseni* (C. L. Gross and P. H. Weston) individual sampled from the *ex situ* germplasm collection was included as an outgroup.

Fresh leaf material was collected, dehydrated using silica beads, and stored at room temperature prior to DNA extraction from single plants. Approximately 0.02 g of dried leaf tissue from each sample was ground in liquid nitrogen and total genomic DNA was extracted using a Plant DNeasy Mini Kit (Qiagen, Germany) according to manufacturer's protocols. DNA concentration was quantified using a Qubit[®] 2.0 Fluorometer dsDNA BR Assay system (Life Technologies, United States) with 2 μ L of each DNA sample. The size and quality of the DNA extracts were also visualized on a 0.8% TAE agarose gel.

Library Preparation and Sequencing

Genomic DNA was normalized to 50 ng/ μ l for library preparation. Sequence libraries for each sample were prepared using an Illumina Nextera XT DNA Library Preparation Kit following manufacturer's instructions (Illumina, United States). Sequence libraries were quantified using a Bioanalyzer 2100 (Agilent, United States). Each *M. integrifolia* sample was barcoded with a unique index and libraries were pooled, and whole genome sequence data was generated using an Illumina HiSeq 2500 instrument at AGRF, Melbourne. Paired end sequence data (2 bp \times 125 bp reads) were produced from pooled, indexed libraries of approximately 300 bp insert size. Sequence data (2 bp \times 125 bp reads) for the unpooled, single *M. janseni* library was generated using a MiSeq instrument at Southern Cross University with the library preparation procedures described above.

Reference Mapping and SNP Calling

Quality control of raw sequence reads was performed using FastQC², and adapter sequences and low quality bases were trimmed using Trimmomatic (Bolger et al., 2014). Reads \geq 75

²<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

TABLE 1 | Origin and source of wild (A) and cultivated (B) *Macadamia integrifolia* accessions.

(A) Wild accessions.						
Code	Germplasm origin	Sample origin	nReads	Coverage	Chlorotype	nSNPs
W01.MB1	Mount Bauple	CREEC Burpengary, QLD	144,435	204	C1.2	125
W02.MB3	Mount Bauple	CTH Alstonville, NSW	261,134	243	C1.1	124
W02.MB5	Mount Bauple	CTH Alstonville, NSW	180,247	169	C1.3	123
W03.MB1	Mount Bauple	CREEC Burpengary, QLD	111,165	155	C1.4	127
W04.MB1	Mount Bauple	CTH Alstonville, NSW	109,728	90	C2.1	0
W04.MB3	Mount Bauple	CTH Alstonville, NSW	43,003	101	C2.5	7
W04.MB4	Mount Bauple	CREEC Burpengary, QLD	141,397	197	C2.5	7
W05.MB5	Mount Bauple	CTH Alstonville, NSW	460,175	645	C1.2	121
W06.MCk3	Marys Creek	CTH Alstonville, NSW	390,181	363	C2.6	7
W06.MCk6	Marys Creek	CTH Alstonville, NSW	142,934	133	C2.6	7
W07.MR1	Mary River	CREEC Burpengary, QLD	40,774	57	C2.2	2
W08.Mo3	Mooloo	CTH Alstonville, NSW	75,256	175	C2.1	0
W08.Mo4	Mooloo	CTH, Alstonville, NSW	453,911	422	C2.1	0
W10.Am2	Amamoor Creek	CTH Alstonville, NSW	347,049	324	C2.4	2
W11.Am6	Amamoor Creek	CTH Alstonville, NSW	222,293	207	C2.4	2
W12.Am3	Amamoor Creek	CTH Alstonville, NSW	122,205	114	C2.3	2
W15.Vi4	Villeneuve	CTH Alstonville, NSW	375,635	351	C3.1	249
W15.Vi5	Villeneuve	CTH Alstonville, NSW	195,513	183	C3.2	251
W16.CP1	Campbells Pocket	MNGC Tiaro, QLD	47,966	112	C3.7	253
W17.UC4	Upper Caboolture	CTH Alstonville, NSW	263,225	245	C3.4	248
W17.UC6	Upper Caboolture	CTH Alstonville, NSW	1,160,658	1086	C3.3	249
W19.Sa1	Samford Valley	CREEC Burpengary, QLD	107,813	150	C3.10	247
W20.Sa2	Samford Valley	CTH Alstonville, NSW	23,356	55	C3.11	254
W21.Sa2	Samford Valley	CTH Alstonville, NSW	147,856	139	C3.5	249
W22.MN7	Mount Nebo	MNGC Tiaro, QLD	62,514	147	C3.6	254
W23.HP1	Holland Park	CTH Alstonville, NSW	115,706	108	C5.5	242
W23.HP4	Holland Park	CTH Alstonville, NSW	143,375	134	C5.4	243
W24.Tin2	Tingalpa	CTH Alstonville, NSW	225,728	210	C5.2	243
W25.MC1	Mount Cotton	CREEC Burpengary, QLD	46,977	65	C5.1	245
W26.MC6	Mount Cotton	CTH Alstonville, NSW	69,342	65	C5.3	246
W27.Bee1	Beenleigh	CREEC Burpengary, QLD	72,425	102	C5.10	243
W28.Or1	Ormeau	CTH Alstonville, NSW	177,786	165	C5.6	243
W28.Or4	Ormeau	CTH Alstonville, NSW	249,693	232	C5.7	241
W30.WV	Willowvale	Same as germplasm	263,981	370	C5.11	231
W29.Or5	Ormeau	CREEC Burpengary	173,190	243	C5.12	240
W31.Co1	Upper Coomera	MNGC Tiaro, QLD	41,874	98	C3.9	253
		Mean	195,729	214.4		157.5

Germplasm origin refers to origin of the original tree (ortet) and sample origin refers to the location of the clone (ramet) from which samples were collected for this study. nReads, number of sequence reads mapped; Coverage, average read coverage; Chlorotype, sub-clade and chloroplast haplotype; nSNP, number of single nucleotide polymorphism differences relative to the reference genome, cultivar HAES 741; CREEC, Caboolture Region Environmental and Education Centre, Burpengary Australia; CTH, Centre for Tropical Horticulture, Alstonville; MNGC, Macadamia National Germplasm Collection.

base pairs (bp) in length with a minimum Q-value of 20 were retained for further analysis. The complete chloroplast genome sequence of *M. integrifolia* cultivar 741 'Mauka' (GenBank Accession No. KF862711) was used as a reference to identify SNP variants. Paired-end reads were mapped to the reference using SOAPaligner (Gu et al., 2013) allowing a maximum of two mismatches per read. Reads with low-quality alignments were identified and filtered out using SAMTOOLS with default parameters (Li et al., 2009). The programs Genome Analysis Toolkit, GATK (DePristo et al., 2011) and Picard Tools³ were

used to optimize alignments by realigning reads around indels and removing duplicate reads following GATK best practices⁴. Following variant calling, the alignment was manually curated to remove low quality (Q < 10) sites. Mapping files (BAM) were used to identify SNPs for each sample in comparison to the reference genome of cultivar 741 using the SNP discovery pipeline SGSautoSNP that was developed for medium coverage resequencing data. In comparison, SAMtools/BCFtools requires extensive filtering to achieve similar true-positive rates of SNP discovery (Lorenz et al., 2012). Individual alignments

³<http://broadinstitute.github.io/picard/>

⁴<https://software.broadinstitute.org/gatk/best-practices/>

TABLE 1 | (B) Cultivated germplasm.

Code	Germplasm origin	Sample origin	nReads	Coverage	Chlorotype	nSNPs
A.Yer	Yeronga, AU	Same as germplasm	30,066	70	C4.1	245
A.WH	Brisbane Botanical Gardens, AU	Same as germplasm	291,521	271	C4.2	235
H.246	HMNC Keauhou orchard, HI	CTH Alstonville, NSW	657,315	921	C2.1	0
H.294	HMNC Nutridge orchard, HI	CTH Alstonville, NSW	449,509	630	C2.1	0
H.333	HMNC Nutridge orchard, HI	CTH Alstonville, NSW	157,290	221	C2.1	0
H.344	HMNC Nutridge orchard, HI	CTH Alstonville, NSW	645,193	905	C2.1	0
H.425	HMNC Keauhou orchard, HI	UCR, CA	140,653	197	C2.1	0
H.508	HMNC Nutridge orchard, HI	USDA Waiakea, HI	53,892	76	C2.1	0
H.660	Deschwanden orchard, HI	CTH Alstonville, NSW	313,874	440	C2.1	0
H.741	Deschwanden orchard, HI	USDA Waiakea, HI	286,819	402	C2.1	0
H.HSp	HSC Kapulena orchard, HI	Honokaa, HI	406,422	570	C2.1	0
H.814	Malamka-Ki UH Exp. Station, HI	CTH Alstonville, NSW	482,659	676	C2.1	0
H.791	Bond Orchard, HI	CTH Alstonville, NSW	219,900	308	C1.5	122
H.Cwy	Coelho Way, Honolulu, HI	Same as germplasm	106,417	99	C2.1	0
H.Purv	Kapulena, HI	Same as germplasm	33,077	47	C2.1	0
H.TPN	Kapulena, HI	Same as germplasm	39,635	93	C2.1	0
H.TPS	Kapulena, HI	Same as germplasm	31,831	74	C2.1	0
H.Wai	Waipio Valley, HI	Same as germplasm	22,041	52	C2.1	0
H.Nut03	HMNC Nutridge orchard, HI	Same as germplasm	36,690	52	C2.1	0
H.Nut07	HMNC Nutridge orchard, HI	Same as germplasm	36,003	51	C2.1	0
H.Nut12	HMNC Nutridge orchard, HI	Same as germplasm	26,481	36	C2.1	0
H.Nut14	HMNC Nutridge orchard, HI	Same as germplasm	47,331	67	C1.6	126
H.Nut15	HMNC Nutridge orchard, HI	Same as germplasm	103,434	146	C2.1	0
C.UCB	UC Berkeley, CA	Same as germplasm	41,670	98	C4.1	247
C.Hei	Coronado, CA	Same as germplasm	13,254	31	C5.8	251
C.Fau	Faulkner Farm, CA	Same as germplasm	24,057	56	C5.9	248
		Mean	180,655	253.4		56.7

Prefix of accession code indicates origin of germplasm. A, Australia; H, Hawaii; C, California; HMNC, Hawaiian Macadamia Nut Company; HSC, Honokaa Sugar Company; UH, University of Hawaii; UCR, University of California Riverside.

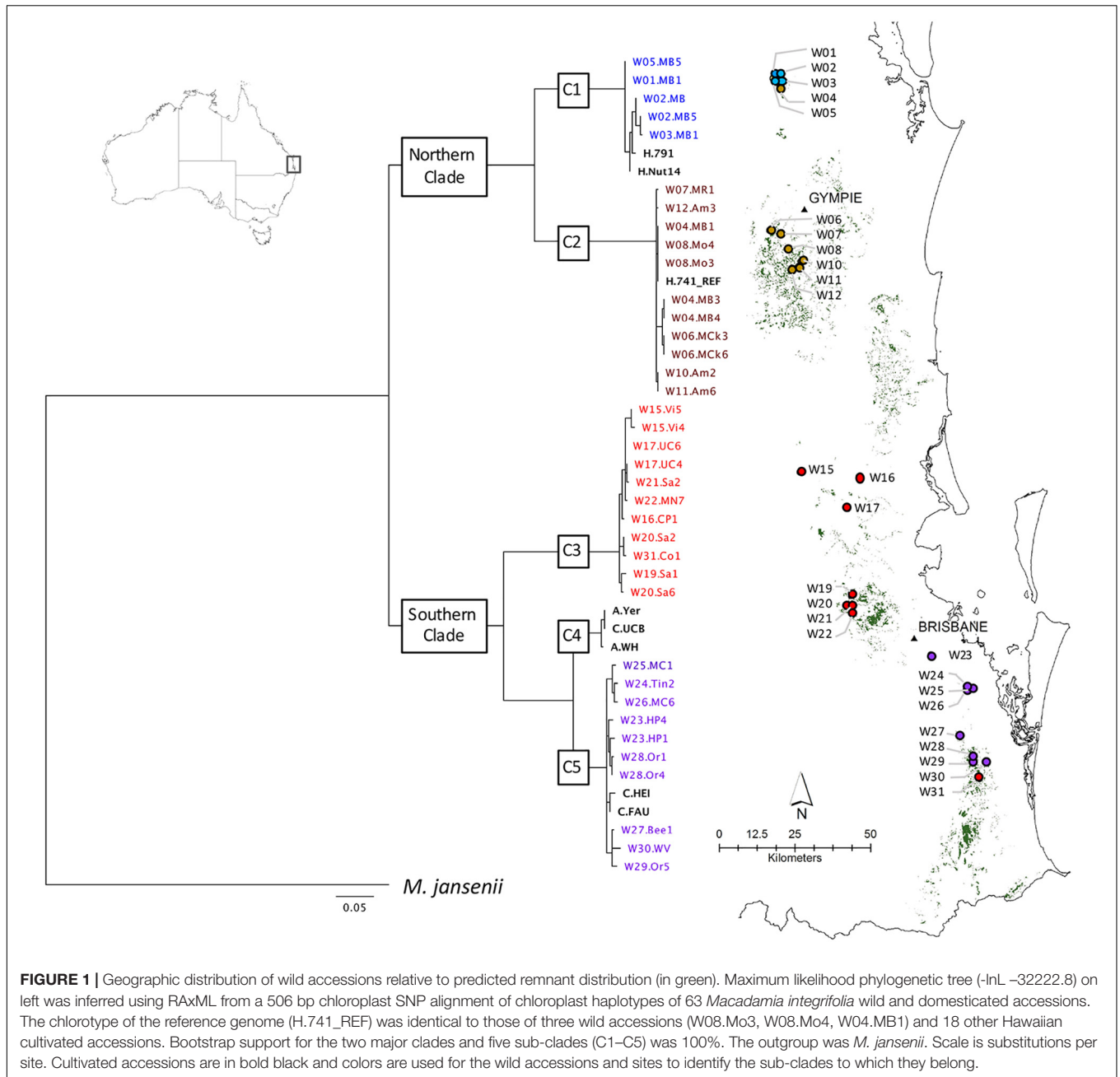
were collated to produce a single variant call format (vcf) table for all samples using SAMTOOLS that was filtered to include only high-quality informative SNP sites with minimum coverage of 10x per sample. The program SnpEff (Cingolani et al., 2012) was used to annotate and predict the effects of SNPs in *M. integrifolia*.

Phylogenetic Analysis

The vcf file with the final set of SNPs was converted into a concatenated sequence alignment of variable positions in fasta format using a custom perl script. Invariant positions were removed and a concatenated sequence alignment of variable positions was produced. *M. janseni* was selected as outgroup because it is geographically isolated so does not naturally hybridize with *M. integrifolia*. The program JModeltest 2 was used to select an optimal substitution model for phylogenetic analysis (Darriba et al., 2012). Maximum likelihood analyses were conducted using Randomize Accelerated Maximum Likelihood RAXML 8.1.2 (Stamatakis, 2014) using raxmlGUI (Silvestro and Michalak, 2012) applying the most-likely substitution model (GTR+G, -lnL 3834.28, γ -shape parameter 99.81). To determine phylogeographic structure and the likely origin of cultivars and other cultivated

germplasm, phylogenetic analysis was conducted independently on alignments of wild accessions and the total dataset including wild and cultivated accessions with gaps treated as missing data. In each case, to determine the optimal phylogeny and assess reliability, analyses implemented 1000 bootstrap replicates and 10 subsequent thorough maximum likelihood (ML) searches. Phylogenetic trees were viewed in FigTree 1.4.3. The relationships between distinct haplotypes were visualized using a statistical parsimony network (Templeton et al., 1992) constructed using TCS 1.21 (Clement et al., 2000).

To examine the relationship between SNP function and the phylogenetic structure, SNP variation was classified according to variation among and within phylogenetic clades and sub-clades. The predicted functional effect of a SNP was compared to its phylogenetic class, and a two-way chi-square analysis of the function-by-geographic structure contingency table of SNP characteristics was undertaken to test the hypothesis that the distribution of SNP function was independent of phylogenetic structure. For this test, sub-clade specific SNPs were collapsed into a single class and non-specific, intragenic, stop-gain, and stop-loss classes were excluded due to low numbers.



RESULTS

Sequencing and Mapping

Raw sequence reads of 64 macadamia accessions were obtained and mapped to the chloroplast genome of *M. integrifolia* cultivar 741 (GenBank Accession No. KF862711). An average of 189,508 reads per *M. integrifolia* accession were mapped to the reference genome. Average read coverage was 214x for wild accessions and 253x for cultivated germplasm samples and ranged from 31 to 1086x per accession (Table 1). For the *M. janseni* accession, 51,112,404 reads mapped to the reference with a mean coverage of 4,820x.

Identification and Analysis of SNP Variation

Following GATK mapping and manual curation, 506 non-redundant SNP sites were identified across the chloroplast genomes of 64 samples including the outgroup *M. janseni*. Concurrent research indicates that the IR regions of the chloroplast genome are highly conserved between *Macadamia* species, for example only five IR single nucleotide polymorphisms (SNPs) were detected between *M. integrifolia* and *M. janseni* (Nock, unpublished data). In this study, all intraspecific *M. integrifolia* SNPs were located in the LSC and SSC single copy regions only. Of these, 407 were variable within *M. integrifolia*

and the average intraspecific single copy region SNP density was 3.8 SNPs per kb (**Figure 2**). Most variants were bi-allelic, however, 12 tri-allelic sites were identified including four within *M. integrifolia*. While the majority of intraspecific variants were located in the LSC region (310, 75.8%), SNP density was greatest in the SSC region (5.3 SNPs per kb, compared to 3.5 SNPs per kb in LSC). SNP variants were distributed across the single copy regions. However, SNP density was elevated in some regions with > 10 SNPs per kb spanning base positions 5–6 kb, 9–11 kb in the LSC and 130–131 kb in the SSC. Alternatively, some sections of the LSC were highly conserved with no SNPs detected within 23–24.5, 37–38.5, 52.5–54, 55.5–57, and 57.8–59.9 kb (**Figure 2**). Of the 407 intraspecific variant sites, 242 (59.5%) were located in non-coding regions and 165 (40.5%) were in exons. Variant sites within exons were located in 48 of 78 genes (61.5%) in the chloroplast single copy regions with most containing a single SNP. Thirteen genes were affected by > 3 variant positions, and the most variable genes were *ycf1* and *ndhF* with 30 and 17 SNP sites respectively (**Table 2**). Based on 506 non-redundant SNP sites, the non-synonymous to synonymous SNP ratio was 1:2. Three exonic variants were nonsense mutations, 97 were missense and 64 were silent. Among the non-synonymous mutations identified, nonsense mutations affected only two genes, *ndhF* (2) and *rpl16* (1), while missense mutations were detected in 36 genes (**Supplementary Table S1**).

Shared Haplotypes

In total, 38 distinct chloroplast haplotypes (chlorotypes) were identified, with one to 257 chloroplast SNP differences between them (**Figure 3**). The reference chlorotype of Hawaiian cultivar HAES 741 (H.741) was shared with two wild accessions sampled from the Mooloo site (W08.Mo3 and W08.Mo4) and one from a Mt Bauple site (W04.MB1). Chlorotypes were identical among samples from the same site (W04.MB3 and W04.MB4; W06.MCk3 and W06.MCk6), across sites within

the Mt Bauple locality (W01.MB1 and W05.MB5) and within the Amamoor locality (W10.AM2 and W11.AM6). However, multiple chlorotypes were also found at other sites from which multiple accessions were sampled (W02, W04, W15, W17, W20, W23, and W28). The reference chlorotype was also identical to those of nine other Hawaiian cultivars (H.508, H.246, H.HSp, H.660, H.814, H.294, H.425, H.344, H.333) and nine other cultivated Hawaiian accessions (H.Nut03, H.Nut07, H.Nut12 and H.Nut15, H.Cwy, H.Purv, H.TPS, H.TPN, H.Wai). Interestingly, a single chlorotype was shared by two cultivated trees: one a backyard tree planted in Brisbane, Australia (A.Yer) and the other planted at the University of California Berkeley, United States (C.UCB). Of 21 Hawaiian accessions included in the study there were three distinct chlorotypes including the reference (cultivar 741), cultivar H.791 and H.Nut14. Cultivar H.791 and H.Nut14 differed from the reference chlorotype at 122 and 126 SNP positions respectively (**Table 1**).

Phylogenetic Analysis

A concatenated sequence alignment of 506 SNP positions was used for phylogenetic analysis. The proportion of gaps was 1.41% and GC content was 63.4%. The best maximum likelihood tree (lnL = -32222.8) produced with the GTR+G model for all accessions shared the same topology as the best tree from analysis of a reduced dataset containing wild accessions only (lnL = -3221.4). Phylogenetic reconstruction revealed five well-supported clades. The tree was rooted with *M. janseni* as the outgroup and there was maximum 100% bootstrap support for each of the two major clades and five sub-clades, C1–C5 (**Figure 1**).

Structure of Wild Populations

There was a clear relationship between the phylogenetic structure and geographic origin of the wild accessions of

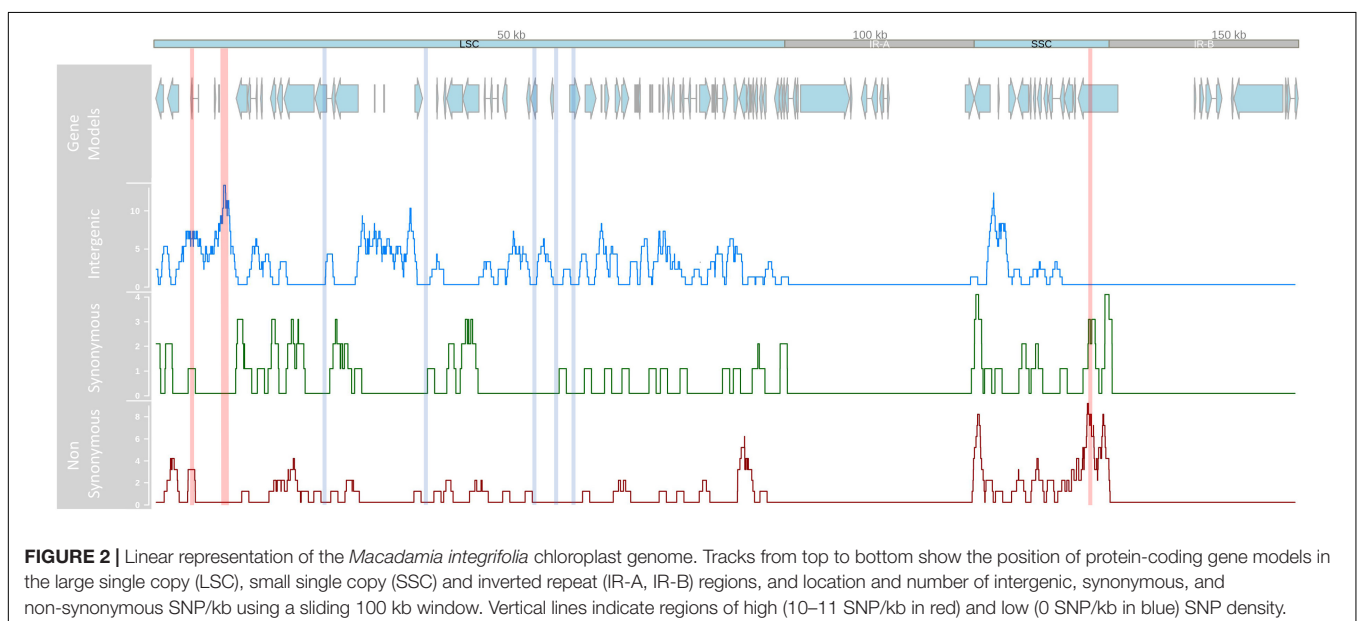


TABLE 2 | Highly variable *Macadamia integrifolia* chloroplast protein coding genes.

Gene	Protein	Length bp	N n	S n	SNP	Density SNP/kb
<i>ycf1</i>	Putative protein TIC214	5538	22	8	30	5.42
<i>ndhF</i>	NAD(P)H-quinone oxidoreductase subunit 5	2220	12	5	17	7.66
<i>rpoC2</i>	DNA-directed RNA polymerase subunit beta"	4185	6	5	11	2.63
<i>rpoB</i>	DNA-directed RNA polymerase subunit beta	3213	3	5	8	2.49
<i>matK</i>	Maturase K	1530	5	2	7	4.58
<i>psaA</i>	Photosystem I P700 chlorophyll a apoprotein A1	2253	2	5	7	3.11
<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	987	5	0	5	5.07
<i>ndhA</i>	NAD(P)H-quinone oxidoreductase subunit 1	2194	3	1	4	1.82
<i>rps16</i>	30S ribosomal protein S16	1090	3	1	4	3.67
<i>ndhD</i>	NAD(P)H-quinone oxidoreductase chain 4	1506	2	2	4	2.66
<i>psaB</i>	Photosystem I P700 chlorophyll a apoprotein A2	2205	2	2	4	1.81
<i>atpA</i>	ATP synthase subunit alpha	1524	1	3	4	2.62
<i>atpI</i>	ATP synthase subunit a	744	1	3	4	5.38

Genes containing four or more SNPs, single nucleotide polymorphisms; bp, base pairs; N, non-synonymous; S, synonymous; kb, kilobase.

M. integrifolia in this study (Figure 1). Two major clades were identified. The northern clade contained all accessions sampled from sites around Amamoor in the Gympie region north to Mt Bauple which is the northern limit of *M. integrifolia*. The southern clade contained all wild accessions from sites from Villeneuve approximately 70 km northwest of Brisbane south to Upper Coomera approximately 50 km southeast of Brisbane. Within the northern clade, sub-clade C1 included accessions from four wild sites within the Mt Bauple region while sub-clade C2 included all accessions from Gympie to Amamoor. However, sub-clade C2 also included trees from a fifth Mt Bauple site (W04.MB). Accessions from wild sites to the south of the Amamoor region belonged to two sub-clades (C3 and C5) of the major southern clade. Sub-clade C3 contained accessions from sites between Amamoor and the Brisbane river, with the exception of a single accession from a site south of the Brisbane river (W31.Co1). Sub-clade 5 contained all other accessions from sites south of the Brisbane River from Holland park (W23.HP) to Willowvale (W30.WV).

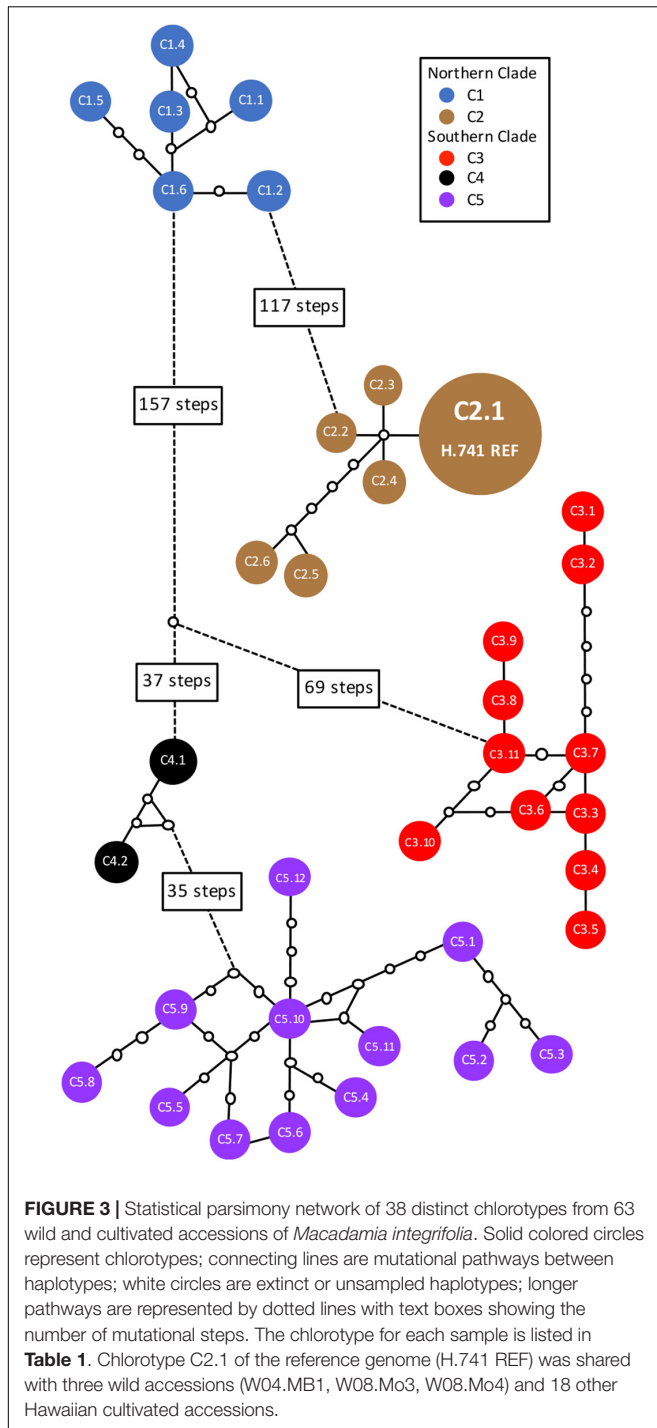
Divergences within sub-clades were very shallow compared to the deeper divergences between the northern and southern clades, and sub-clades C1–C5 (Figure 1). Mutational steps separating the 13 northern and 25 southern haplotypes ranged from a minimum of 195 to 256, and within sub-clades from one to six to 16 steps (Figure 3). Further evidence of the phylogenetic structure of chloroplast variation in *M. integrifolia* was provided by the distribution of SNP variation among phylogenetic clades. Of 407 SNP variants in total, the majority (84.3%) were diagnostic for clades or sub-clades. There were 106 fixed differences between the northern and southern clades while 237 sub-clade specific SNPs were fixed (Table 3). A chi-square test for independence between clade-level phylogenetic structure and SNP annotation class (non-coding, synonymous, and missense) was not significant ($P = 0.486$). Most of the SNP variants were located in non-coding regions (244) or were synonymous mutations (64) and predicted to have no effect on gene function.

Structure of Cultivated Accessions

The chlorotype of the reference accession (H.741), and the other 18 Hawaiian accessions included in this study were identical to the independently assembled reference genome of cultivar 741 (Nock et al., 2014), and were located within sub-clade C2 of the northern clade (Figure 1). The two Hawaiian accessions with unique chlorotypes different to the reference genome (H.791 and H.Nut14) belonged to the most northern sub-clade C1 that also included all accessions from four sites W01, W02, W03, and W05 at Mt. Bauple. None of the accessions from cultivated germplasm were associated with sub-clade C3. Two closely related Californian accessions (C.Fau and C.Hei) belonged to the southernmost sub-clade C5, and differed by a minimum of four substitutions from the wild accessions within this sub-clade. The southern sub-clade C4 was exceptional in that it contained only cultivated accessions including the Walter Hill tree (A. WH) and two other planted trees that shared identical chlorotypes (A.Yer and C.UCB).

DISCUSSION

This study demonstrates that phylogeographically structured intraspecific chloroplast sequence variation can be used to locate the wild sources of cultivated macadamia germplasm. For crops with long histories of domestication and multiple origins, their geographic origin often remains unresolved (Meyer et al., 2012). In contrast, commercial macadamia production developed only recently and wild populations may have been relatively undisturbed prior to the 19th century. Global macadamia production is primarily based on grafted cultivars selected through breeding programs in Hawaii that were released less than 100 years ago, and may be only one to three generations from the wild (Hardner, 2016). The results of this study suggest that these cultivars were derived from a narrow seed pool and provide evidence for a genetic bottleneck in the maternal lineage of this recently domesticated nut crop.



Phylogeographic Structure of *Macadamia integrifolia*

Phylogenetic analysis of chloroplast genomic variation revealed a latitudinal population structure of wild *M. integrifolia* germplasm, suggesting long-term regional isolation of maternal lineages (**Figure 1**). The deep divergence between northern and southern clades is indicative of an historical barrier to seed dispersal north of Brisbane, between the Brisbane and Mary River

catchments. This finding is concordant with the only previous intraspecific genetic analysis of *M. integrifolia*, using nuclear randomly amplified DNA fingerprinting (Peace, 2002). The two major clades of *M. integrifolia* identified in this study are located within two separate subtropical refugia, or centers of endemism, defined by Weber et al. (2014). In addition, some of the suitable habitat in the region dividing the northern and southern clades is occupied by *M. ternifolia* (F.Muell) and hybrid populations (Costello et al., 2009; Hardner et al., 2009). These factors are likely contributors to the divergence between northern and southern *M. integrifolia* populations. Limited comparative chloroplast genome data for other Proteaceae taxa precludes reliable dating of intraspecific *M. integrifolia* divergences at this time. However, a crown age, for the subtropical genus *Macadamia* and its most recent common ancestor, of approximately 7 Mya was inferred from a fossil-calibrated phylogeny of the tribe Macadamieae taxa based on six chloroplast and nuclear genes (Mast et al., 2008). This period was coincident with the late Miocene contraction and fragmentation of rain forest habitat and aridification of much of the Australian continent (Byrne et al., 2008). During the Pliocene (approximately 5.3 to 1.8 mya) subtropical rain forest is thought to have persisted only on some regions of the Great Dividing Range and east coast with subsequent expansion and contraction during glacial and interglacial periods during the Quaternary (Byrne et al., 2011; Weber et al., 2014).

Further geographic structuring of genetic variation was detected within each of the major northern and southern clades. In the north, sites from Mt Bauple (C1) and Gympie (C2) regions formed two distinct clades. In the south, two clades including trees from sites north-west of Brisbane (C3) and south of Brisbane to the Gold Coast (C5) were separated by the Brisbane River Valley. Extensive evidence supports the existence of multiple biogeographic barriers in eastern Australia that led to vicariance events in rain forest restricted flora and fauna, including the speciation of *Macadamia* (reviewed in Weber et al., 2014; Bryant and Krosch, 2016). Spatial habitat modeling has been used to predict historical and remnant *M. integrifolia* habitat (Powell et al., 2010, 2014). Fragmentation and numerous gaps in the distribution of suitable habitat were identified particularly in the region separating the northern and southern clades, and the Brisbane River Valley separating the southern subclades (**Figure 1**). Our findings suggest that genetic divergence within *M. integrifolia* was the consequence of multiple barriers to seed dispersal between the lowland coastal ranges of subtropical eastern Australia. There was limited evidence for admixture between sites within the northern and southern regions, supporting the assumption that most of the sites sampled were remnant vegetation. Individuals from one Mount Bauple site (W04) were more closely related to those from the Gympie region than to those of proximally located sites suggesting that this site may contain translocated germplasm (**Figure 3**). Similarly, one tree from Upper Coomera (W31) on the Gold Coast belonged to a sub-clade (C3) that otherwise included only wild trees from sites north of Brisbane. The geographic-genetic discordance of individuals from these two presumed wild sites could be due to the long distance, possibly human-mediated, translocation of seed. Increased sampling and further research is

TABLE 3 | Distribution of *Macadamia integrifolia* chloroplast SNP variation by annotation type and phylogenetic structure.

Phylogenetic class	SNP annotation type						Total
	Non-coding		Synonymous	Non-synonymous			
	Intergenic	Intragenic		Missense	Stop gain	Stop loss	
Clade specific	54	5	19	27	1		106
Sub-clade specific							
C1	28	3	10	8			49
C2	28	3	11	17			59
C3	33		10	13			56
C4	11		1	3			15
C4+C5	17	1	3	4			25
C5	19		5	9			33
Within sub-clade	16	1	4	12			33
Non-specific	23	1	1	4		2	31
Total	229	14	64	97	1	2	407

Clade specific, polymorphisms fixed within the northern and southern clades. Sub-clade specific, fixed polymorphisms within each sub-clade (including C4+C5). Within sub-clade, variable within a single sub-clade only. Non-specific, unrelated to phylogenetic structure.

needed to understand the extent of historical and more recent human-mediated dispersal in this species.

Wild Origins of Macadamia Domestication

The high chloroplast variability and geographic structure of this variation support the use of chloroplast sequence data to identify wild origins of cultivated germplasm in macadamia. This study sampled three distinct maternal lineages, or chlorotypes, from the Hawaiian germplasm. All first generation cultivars that were selected from seedling orchards established from the early 1920s to the mid 1930s shared a single chlorotype. Subsequent Hawaiian cultivars and selections were from progeny of these selections (predominantly H.246) or from germplasm that was introduced into Hawaii in the 1950s (Hardner, 2016). This chlorotype was shared by all cultivars and cultivated germplasm sampled from Hawaii with the exception of H.791 and H.Nut14. The same chlorotype was also present in three wild trees from sites at Mooloo and Mt. Bauple suggesting that the maternal lineage of almost all Hawaiian cultivars may trace back to one wild site, and perhaps even a single tree within a site. The Mt Bauple region and Mooloo valley, south west of Gympie in the north of the *M. integrifolia* distribution are still relatively undisturbed compared to other parts of the predicted pre-colonization distribution of macadamia. It is possible that the original trees from which seed was collected and taken to Hawaii may still be alive today. The differentiation of two other Hawaiian accessions (H.791 and H.Nut14) demonstrates some diversity in the maternal lineages of Hawaiian germplasm. Their likely ancestral origin is Mt Bauple given that their chlorotypes belong to a clade that otherwise includes only individuals from this region. Although closely related, the chlorotypes of H.791 and H.Nut14 differ by three mutational steps from each other so must have been derived from seed from different wild trees (Figure 3). These results agree with an earlier study suggesting that the Hawaiian cultivars originated from the

north of the *M. integrifolia* distribution (Peace, 2002). Previous studies have examined genetic relationships among macadamia cultivars. Moderate variation was identified among Hawaiian cultivars using 16 allozyme loci although most alleles were shared (Aradhya et al., 1998). Subsequent analyses based on dominant AFLP and DNA RAF markers placed Hawaiian cultivars in separate but closely related *M. integrifolia* clusters (Steiger et al., 2003; Peace et al., 2005).

Contribution of Reported Introductions to Hawaiian Domesticated Germplasm

Historical records on the development of the Hawaiian macadamia industry may provide some evidence about the contribution to the Hawaiian germplasm of different introductions and the wild origin of these introductions.

The Purvis Introduction

There is limited information on the possible origin of the first introduction of macadamia by W. H. Purvis into Hawaii. Hardner (2016) suggests that Purvis may have obtained germplasm in England in late 1882. He personally records taking Wardian cases of plants with him on his departure to return to Hawaii although the source of this material is unknown. Alternatively, seed (or plants) may have been obtained when he stopped in Sydney for a month on the same voyage. Some reports suggest that *M. ternifolia* was also included in the Purvis introduction. The samples H.TPN, H.TPS, and H.Purvis are believed to represent the original Purvis plantings (Wagner-Wright, 1995; Hardner, 2016). The trees H.TPS and H.TPN grow in Kapulena, about 4.5 km from Purvis' House in Kukuhaele. Honokaa Special (H.HSp) and H.Wai are reportedly from the Honokaa Sugar Company seedling orchard, while cultivar H.791 was reportedly selected from the Bond seedling orchard which was planted around 1922 with seedlings possibly supplied by Honokaa Seedling orchard. Samples included in this study that are presumed to represent

the Purvis introduction all shared the same chlorotype with the exception of cultivar H.791.

The Jordan Introduction

The second documented introduction of macadamia to Hawaii was reportedly *M. integrifolia* seed shipped from Queensland in 1892 (Hamilton and Storey, 1956; Shigeura and Ooka, 1984). An old tree sampled in Hawaii (H.Cwy) is believed to be closely related to the Jordan introduction based on: (i) the age of the tree inferred from its large size, and (ii) the proximity, less than 80 m, to the last known survivor of the original six trees of the Jordan introduction (Shigeura and Ooka, 1984). Cultivars H.294, H.333, H.344, H.508 and other samples from the Nutridge orchard (H.Nut, **Table 1B**) are also thought to represent the Jordan introduction although this orchard may also have had minor contributions from other (*M. tetraphylla* and Purvis) sources. The Keauhou orchard, from which H.246 and H.425 were selected, was reportedly planted with germplasm that was only sourced from the Jordan introduction. In addition, Hardner (2016) suggests that the Deschwanden orchard, which produced H.660 and H.741, was established with germplasm collected from the Nutridge orchard. All 14 samples that are presumably representative of the Jordan introduction also shared the reference H.741 chlorotype with the exception of one (H.Nut14) from the Nutridge orchard.

Results from this study suggest that both the Jordan and Purvis introductions were derived from seed from at least three trees from northern *M. integrifolia* populations – the Jordan introduction from a lineage at Mooloo or Mt Bauple, and the Purvis introduction from a different Mt Bauple lineage. It has been suggested that a living tree at Willowvale in southeast Queensland (W30.WV) was the source of seed for the Jordan introduction (Lowndes, 1996). However, the chlorotype of this tree was most closely related to other samples in the most southern sub-clade C5 indicating that neither this tree, nor indeed any wild tree from this region, was a maternal parent of the Hawaiian germplasm included in this study. That the same maternal lineage is shared among the majority of Hawaiian samples in this study suggests that they were either derived from (i) the same introduction, or (ii) more than one introduction from the same region and the same maternal lineage. The other two maternal lineages identified include one cultivar (H.791) with *M. ternifolia* content (Peace et al., 2005).

The Genetics of Extinct Wild Populations May Be Preserved in Cultivated Trees

Of interest is the sub-clade (C4) that includes two Australian and one Californian cultivated sample but included no wild accessions. This clade contains one of the oldest recorded cultivated macadamia planted by Walter Hill in the Brisbane Botanic Gardens, Brisbane (A.WH). Unfortunately, his records were destroyed in a flood so no information of the origin of this tree is available. This sub-clade also included a tree from a suburban backyard south of Brisbane (A.Yer) and another from the University of California (C.UCB). The distinctiveness of this sub-clade suggests that these trees may represent an extinct

population, or a separate source of wild germplasm that was not sampled in this study. The phylogeographic structure of chloroplast variation suggests that these trees most likely trace back to a population south of the Brisbane River. Brisbane was settled in 1824 as a penal colony and much of the urban area was subsequently cleared due to development. It may be that wild populations existed in the region prior to the turn of the 20th century when the population of Brisbane was small (135,000) compared to the present population of over two million. Although natural populations of macadamia have likely been lost since European occupation, our results suggest that planted trees in parks, gardens, and private backyards may represent a source of unique germplasm for future breeding.

The other two cultivated samples from California (C.Fau, C.Hei) trace back to wild sites associated with the southernmost sub-clade C5, and are most closely related to trees from sites south of Brisbane including Holland Park, Beenleigh and Ormeau. While their chlorotypes are closely related, there are no reports on the relationship between these trees.

The Chloroplast Genome of *Macadamia integrifolia* Is Highly Variable

The chloroplast genome has been used extensively to resolve evolutionary relationships among plant species, and the capacity to detect variation has markedly improved since the advent of NGS (Chase et al., 1993; Parks et al., 2009; Moore et al., 2010; Soltis et al., 2011; Ruhfel et al., 2014). However, fewer studies have used whole chloroplast genome data to examine intraspecific diversity. Here we have sequenced the chloroplast genome of 64 accessions sampled from remnant wild and cultivated *Macadamia integrifolia* germplasm. Macadamia belongs to the Proteaceae, a large family of over 1700 species, spanning the remnant landmasses of Gondwana including Southern Africa, South America and New Zealand, and contains other species that are valued for food and floriculture. However, the first genomic data for the family became available only recently and we are unaware of any other phylogenomic study in the Proteaceae.

In total, 407 intraspecific polymorphisms and 38 distinct haplotypes were detected among 63 accessions with an average coverage of over 200x (**Table 1**). Chloroplast diversity in *M. integrifolia* is relatively high compared to that reported for other plant species including *Jacobaea vulgaris* (32 SNPs, 17 individuals), the model grass plant *Brachypodium distachyon* (298 SNPs, 32 haplotypes, 53 individuals), rapeseed *Brassica napus* (294 SNPs, 488 individuals) and Australian rain forest trees (6 to 240 SNPs per species, 12 individuals of 12 diverse species) (Doorduyn et al., 2011; Van der Merwe et al., 2014; Qiao et al., 2016; Sancho et al., 2017). Relatively high genetic diversity is likely a consequence of the long-term persistence of genetically distinct populations in multiple stable subtropical rain forest refugia of eastern Australia through periods of historical climate variability (Weber et al., 2014; Rossetto et al., 2015). *M. integrifolia* is distributed over an approximately

250 km latitudinal range and is restricted to lowland subtropical rain forest on the coastal ranges of south east Queensland (Hardner et al., 2009; Powell et al., 2010, 2014). The large, hard-shelled seeds are thought to be dispersed by *Rattus rattus*, native rodents, gravity, and water (Pisanu, 2001; Neal et al., 2010; O'Connor et al., 2015). Despite an estimated 63% habitat loss and fragmentation due to land clearing, spatial habitat modeling indicates that a network of suitable *M. integrifolia* habitat persists (Powell et al., 2010).

Although the majority of variants were located in the LSC region of the chloroplast genome, SNP density was highest in the SSC region (5.26 SNPs per kb, compared to 3.52 in LSC). This finding has been reported for other plants including *Sesamum indicum* (Zhang et al., 2013) and *Panax ginseng* (Zhao et al., 2015) and is noteworthy because most universal primers and PCR-based intraspecific chloroplast studies assess variation only in the LSC. Regions of particularly high variability in *M. integrifolia* included the LSC intergenic spacer regions *trnQ-rps16* and *trnS-trnG-trnG*. Both have been identified as short, variable, and underutilized regions of the angiosperm chloroplast genome suitable for intraspecific phylogenetic studies (Daniell et al., 2006; Shaw et al., 2007). Within the SSC, the most variable regions included two genes, *ycf1* and *ndhF* (Table 2). *Ycf1* is the most rapidly evolving chloroplast gene (Dong et al., 2015) although its function across plant taxa remains unresolved (de Vries et al., 2015; Nakai, 2015; Bölter and Soll, 2017). The *ndhF* gene encodes a subunit of an NADH-specific dehydrogenase complex involved in photosynthetic electron transport (Yamori et al., 2016). Elevated sequence diversity and loss of function in chloroplast *ndhF* have been reported for a wide range of photosynthetic plant taxa (Wakasugi et al., 1994; Kim and Jansen, 1995; Kim and Chase, 2017). Our results are concordant with a recent phylogenetic analysis of 34 chloroplast genomes from *Citrus* and related species which found *ycf1*, *rpoC2*, *ndhF*, and *matK* to be the most variable chloroplast genes. There was evidence for positive selection of *ndhF* and *matK* exclusively in the Australian lineage (*Microcitrus* and *Eremocitrus*) suggesting that these genes may be involved in adaptation to contrasting climatic conditions (Carbonell-Caballero et al., 2015; Daniell et al., 2016).

CONCLUSION

In this study, geographically structured variation of the *M. integrifolia* chloroplast genome was used to identify the wild origin and a maternal bottleneck in the Hawaiian cultivars that are the basis of the world macadamia industry. In addition, it appears that genetic diversity has been lost in the wild since European colonization, although some of this may be captured in cultivated trees. Comparison of chloroplast variation with that of the nuclear genome could test the hypothesis that the seed used to develop most of the Hawaiian cultivars was collected from a single tree, and will add greater insight into the genetics of the genus and crop.

DATA AVAILABILITY

The datasets analyzed for this study can be found in the European Variation Archive EMBL-EBI (Project: PRJEB2832 Analyses: ERZ683764, <https://www.ebi.ac.uk/ena/data/view/PRJEB28321>).

AUTHOR CONTRIBUTIONS

CN prepared and drafted the manuscript, contributed to the study design, data analysis and interpretation and co-supervised AAT. CH established the *ex situ* conservation trials from which the wild germplasm used in this study was sampled, conceived and designed the study, collected Hawaiian, Californian and Australian cultivated samples, supervised AAT, and collaborated with CN to develop early versions of the manuscript. JM and DE managed data and performed bioinformatic analyses. AAT conducted initial laboratory work and analysis during a B.Sc. (Hons) research project. SH contributed to laboratory work, library preparation and sequencing. JP designed and supervised the original collection of wild germplasm from remnant populations. JB contributed to the data collection, analysis and interpretation and co-supervised AAT. All authors reviewed draft the manuscripts.

FUNDING

This work was supported by Horticulture Innovation Australia, Ltd. using the macadamia research and development levy and contributions from the Australian Government. A Churchill Fellowship Trust provided support to CH for travel to Hawaii to research the domestication history of macadamia germplasm in Hawaii.

ACKNOWLEDGMENTS

The authors thank Steve Falconer for the original collection of wild germplasm, Hidden Valley Plantations for propagation of germplasm, CSIRO for planting *ex situ* germplasm trials, and CSIRO, Queensland Department of Agriculture and Fisheries, the University of Queensland and NSW Department of Primary Industry, Fraser Coast Regional Council for maintenance of the trials. The map was drawn by Melissa Walker following data and methods from Powell et al. (2010). The authors also thank the landholders for access to their properties for sample collection and to Miguel Villamil Castro for assistance with DNA extraction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00334/full#supplementary-material>

REFERENCES

- Aradhya, M. K., Yee, L. K., Zee, F. T., and Manshardt, R. M. (1998). Genetic variability in Macadamia. *Genet. Resour. Crop Evol.* 45, 19–32. doi: 10.1023/A:1008634103954
- Bock, D. G., Kane, N. C., Ebert, D. P., and Rieseberg, L. H. (2014). Genome skimming reveals the origin of the Jerusalem artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytol.* 201, 1021–1030. doi: 10.1111/nph.12560
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bölter, B., and Soll, J. (2017). Ycf1/Tic214 is not essential for the accumulation of plastid proteins. *Mol. Plant* 10, 219–221. doi: 10.1016/j.molp.2016.10.012
- Brozynska, M., Furtado, A., and Henry, R. J. (2016). Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotech. J.* 14, 1070–1085. doi: 10.1111/pbi.12454
- Bryant, L. M., and Krosch, M. N. (2016). Lines in the land: a review of evidence for eastern Australia's major biogeographical barriers to closed forest taxa. *Biol. J. Linn. Soc.* 119, 238–264. doi: 10.1111/bij.12821
- Burger, J. C., Chapman, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95, 113–122. doi: 10.3732/ajb.95.2.113
- Byrne, M., Steane, D. A., Joseph, L., Yeates, D. K., Jordan, G. J., Crayn, D., et al. (2011). Decline of a biome: evolution, contraction, fragmentation, extinction and invasion of the Australian mesic zone biota. *J. Biogeog.* 38, 1635–1656. doi: 10.1111/j.1365-2699.2011.02535.x
- Byrne, M., Yeates, D., Joseph, L., Kearney, M., Bowler, J., Williams, M., et al. (2008). Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Mol. Ecol.* 17, 4398–4417. doi: 10.1111/j.1365-294X.2008.03899.x
- Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol. Biol. Evol.* 32, 2015–2035. doi: 10.1093/molbev/msv082
- Carpenter, R. J., Macphail, M. K., Jordan, G. J., and Hill, R. S. (2015). Fossil evidence for open, Proteaceae-dominated heathlands and fire in the late Cretaceous of Australia. *Am. J. Bot.* 102, 2092–2107. doi: 10.3732/ajb.1500343
- Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., et al. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Mo. Bot. Gard.* 80, 528–580. doi: 10.2307/2399846
- Chen, Y. H., Shapiro, L. R., Benrey, B., and Cibrián-Jaramillo, A. (2017). Back to the origin: in situ studies are needed to understand selection during crop diversification. *Front. Ecology Evol.* 5:125. doi: 10.3389/fevo.2017.00125
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Clement, M., Posada, D., and Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9, 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x
- Costello, G., Gregory, M., and Donati, P. (2009). *Southern Macadamia Species Recovery Plan 2008–2012*. Canberra: Department of the Environment and Heritage, 37.
- Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* 36:e122. doi: 10.1093/nar/gkn502
- Daniell, H., Lee, S.-B., Grevich, J., Saski, C., Quesada-Vargas, T., Guda, C., et al. (2006). Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* 112, 1503. doi: 10.1007/s00122-006-0254-x
- Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772–772. doi: 10.1038/nmeth.2109
- de Vries, J., Sousa, F. L., Bölter, B., Soll, J., and Gould, S. B. (2015). YCF1: a green TIC? *Plant Cell* 27, 1827–1833. doi: 10.1105/tpc.114.135541
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends Plant. Sc.* 20, 525–527. doi: 10.1016/j.tplants.2015.06.012
- Doebly, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell* 127, 1309–1321. doi: 10.1016/j.cell.2006.12.006
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5:8348. doi: 10.1038/srep08348
- Doorduyn, L., Gravendeel, B., Lammers, Y., Ariyurek, Y., Chin-A-Woeng, T., and Vrieling, K. (2011). The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res.* 18, 93–105. doi: 10.1093/dnares/dsr002
- Fuller, D. Q., Willcox, G., and Allaby, R. G. (2011). Early agricultural pathways: moving outside the 'core area' hypothesis in Southwest Asia. *J. Exp. Bot.* 63, 617–633. doi: 10.1093/jxb/err307
- Gray, M. W., and Doolittle, W. F. (1982). Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* 46, 1–42.
- Gross, C. (1995). "Macadamia," in *Flora of Australia Vol. 16. Elaeagnaceae, Proteaceae I*. ed. P. McCarthy (Melbourne: CSIRO), 419–425
- Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for short reads alignment. *Curr. Protoc. Bioinform.* 44, 1–17. doi: 10.1002/0471250953.bi1111s44
- Hamilton, M. (1999). Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Mol. Ecol.* 8, 521–523.
- Hamilton, R. A., and Storey, W. B. (1956). Macadamia nut production in the Hawaiian Islands. *Econ. Bot.* 10, 92–100. doi: 10.1007/BF02985321
- Hardner, C. (2016). Macadamia domestication in Hawai'i. *Genet. Resour. Crop Evol.* 63, 1411–1430. doi: 10.1007/s10722-015-0328-1
- Hardner, C., Pisanu, P., and Boyton, S. (2004). *National Macadamia Conservation Program, MC99029 Report*. Brisbane: CSIRO.
- Hardner, C. M., Peace, C., Lowe, A. J., Neal, J., Pisanu, P., Powell, M., et al. (2009). "Genetic resources and domestication of macadamia," in *Horticultural Reviews*, ed. J. Janick Hoboken (Hoboken, NJ: John Wiley & Sons), 1–125. doi: 10.1002/9780470593776.ch1
- Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., et al. (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol. Biol. Evol.* 24, 1506–1517. doi: 10.1093/molbev/msm077
- Izan, S., Esselink, D., Visser, R. G., Smulders, M. J., and Borm, T. (2017). De novo assembly of complete chloroplast genomes from non-model species based on a K-mer frequency-based selection of chloroplast reads from total DNA sequences. *Front. Plant Sci.* 8:1271. doi: 10.3389/fpls.2017.01271
- Kim, H. T., and Chase, M. W. (2017). Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PLoS One* 12:e0187318. doi: 10.1371/journal.pone.0187318
- Kim, K. J., and Jansen, R. K. (1995). *ndhf* sequence evolution and the major clades in the sunflower family. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10379–10383. doi: 10.1073/pnas.92.22.10379
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lorenc, M. T., Hayashi, S., Stiller, J., Lee, H., Manoli, S., Ruperao, P., et al. (2012). Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. *Biology* 1, 370–382. doi: 10.3390/biology1020370
- Lowndes, A. G. (1966). Two historic macadamia trees in Australia. *Calif. Macadamia Soc.* 12, 29–32.

- Luo, Y., Reid, R., Freese, D., Li, C., Watkins, J., Shi, H., et al. (2017). Salt tolerance response revealed by RNA-Seq in a diploid halophytic wild relative of sweet potato. *Sci. Rep.* 7:9624. doi: 10.1038/s41598-017-09241-x
- 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 diaspora evolution in *Macadamia* and relatives (tribe Macadamieae, Proteaceae). *Am. J. Bot.* 95, 843–870. doi: 10.3732/ajb.0700006
- McConachie, I. (1980). The macadamia story. *Yearb. Calif. Macad. Soc.* 26, 41–47.
- McPherson, H., Van der Merwe, M., Delaney, S. K., Edwards, M. A., Henry, R. J., McIntosh, E., et al. (2013). Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. *BMC Ecol.* 13:8. doi: 10.1186/1472-6785-13-8
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- Miller, A. J., and Gross, B. L. (2011). From forest to field: perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522
- Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4623–4628. doi: 10.1073/pnas.0907801107
- Nakai, M. (2015). YCF1: a green TIC: response to the de Vries et al. Commentary. *Plant Cell* 27, 1834–1838. doi: 10.1105/tpc.15.00363
- Neal, J. M., Hardner, C. M., and Gross, C. L. (2010). Population demography and fecundity do not decline with habitat fragmentation in the rainforest tree *Macadamia integrifolia* (Proteaceae). *Biol. Conserv.* 143, 2591–2600. doi: 10.1016/j.biocon.2010.06.029
- Nikiforova, S. V., Cavalieri, D., Velasco, R., and Goremykin, V. (2013). Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Mol. Biol. Evol.* 30, 1751–1760. doi: 10.1093/molbev/mst092
- Nock, C. J., Baten, A., Barkla, B. J., Furtado, A., Henry, R. J., and King, G. J. (2016). Genome and transcriptome sequencing confirms the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* 17:937. doi: 10.1186/s12864-016-3272-3
- Nock, C. J., Baten, A., and King, G. J. (2014). Complete chloroplast genome of *Macadamia integrifolia* confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. *BMC Genomics* 15(Suppl. 9):S13. doi: 10.1186/1471-2164-15-S9-S13
- Nock, C. J., Waters, D. L., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotech. J.* 9, 328–333. doi: 10.1111/j.1467-7652.2010.00558.x
- 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. Biol. Conserv.* 191, 504–511. doi: 10.1016/j.biocon.2015.08.001
- Palmer, J. D. (1991). "Plastid chromosomes: structure and evolution," in *The Molecular Biology of Plastids*, eds L. Bogorad and I. K. Vasil (San Diego, CA: Academic Press), 5–53. doi: 10.1016/B978-0-12-715007-9.50009-8
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Peace, C., Allan, P., Vithanage, V., Turnbull, C., and Carroll, B. (2005). Genetic relationships amongst macadamia varieties grown in South Africa as assessed by RFLP markers. *S. Afr. J. Plant Soil* 22, 71–75. doi: 10.1080/02571862.2005.10634684
- Peace, C. P. (2002). *Genetic Characterisation of Macadamia With DNA Markers*. Ph.D. Thesis. [St Lucia]: University of Queensland.
- Pisanu, P. C. (2001). *Survivorship of the Threatened Subtropical Rainforest Tree Macadamia Tetraphylla L. Johnson (Proteaceae) in Small Habitat Fragments*. Ph.D. thesis, University of New England, Armidale.
- Powell, M., Accad, A., Austin, M. P., Choy, S. L., Williams, K. J., and Shapcott, A. (2010). Predicting loss and fragmentation of habitat of the vulnerable subtropical rainforest tree *Macadamia integrifolia* with models developed from compiled ecological data. *Biol. Cons.* 143, 1385–1396. doi: 10.1016/j.biocon.2010.03.013
- Powell, M., Accad, A., and Shapcott, A. (2014). Where they are, why they are there, and where they are going: using niche models to assess impacts of disturbance on the distribution of three endemic rare subtropical rainforest trees of *Macadamia* (Proteaceae) species. *Aust. J. Bot.* 62, 322–334. doi: 10.1071/BT14056
- Provan, J., Powell, W., and Hollingsworth, P. M. (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* 16, 142–147. doi: 10.1016/S0169-5347(00)02097-8
- Qiao, J., Cai, M., Yan, G., Wang, N., Li, F., Chen, B., et al. (2016). High-throughput multiplex cpDNA resequencing clarifies the genetic diversity and genetic relationships among *Brassica napus*, *Brassica rapa* and *Brassica oleracea*. *Plant Biotech. J.* 14, 409–418. doi: 10.1111/pbi.12395
- Rossetto, M., McPherson, H., Siow, J., Kooyman, R., Van der Merwe, M., and Wilson, P. D. (2015). Where did all the trees come from? A novel multispecies approach reveals the impacts of biogeographical history and functional diversity on rain forest assembly. *J. Biogeog.* 42, 2172–2186. doi: 10.1111/jbi.12571
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., and Burleigh, J. G. (2014). From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23
- Sancho, R., Cantalapiedra, C. P., López-Alvarez, D., Gordon, S. P., Vogel, J. P., Catalán, P., et al. (2017). Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* 218, 1631–1644. doi: 10.1111/nph.14926
- Sauquet, H., Weston, P. H., Anderson, C. L., Barker, N. P., Cantrill, D. J., Mast, A. R., et al. (2009). Contrasted patterns of hyperdiversification in Mediterranean hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 106, 221–225. doi: 10.1073/pnas.0805607106
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Schroeder, C. A. (1954). Report of the subtropical fruit varieties committee. *Yearb. Calif. Macad. Soc.* 38, 53–61.
- Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275
- Shigeura, G. T., and Ooka, H. (1984). *Macadamia Nuts in Hawaii: History and Production*. Hawaii: University of Hawaii, 91.
- Silvestro, D., and Michalak, I. (2012). raxmlGUI: a graphical front-end for RAxML. *Org. Divers. Evol.* 12, 335–337. doi: 10.1007/s13127-011-0056-0
- Smith, L. S. (1956). New species and notes on Queensland plants. *P. Roy. Soc. Qld.* 67, 50–56.
- Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* 98, 704–730. doi: 10.3732/ajb.1000404
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Steiger, D. L., Moore, P. H., Zee, F., Liu, Z., and Ming, R. (2003). Genetic relationships of macadamia cultivars and species revealed by AFLP markers. *Euphytica* 132, 269–277. doi: 10.1023/A:1025025522276
- Storey, W. B. (1977). History of the macadamia nut. *Yearb. Calif. Macad. Soc.* 23, 35–43.
- Straub, S. C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., and Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Am. J. Bot.* 99, 349–364. doi: 10.3732/ajb.1100335

- Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109. doi: 10.1007/BF00037152
- Templeton, A. R., Crandall, K. A., and Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132, 619–633.
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135. doi: 10.1038/nrg1271
- Trask, E. E. (1962). Historical macadamia trees of California. *Yearb. Calif. Macad. Soc.* 8:48.
- Van der Merwe, M., McPherson, H., Siow, J., and Rossetto, M. (2014). Next-Gen phylogeography of rainforest trees: exploring landscape-level cpDNA variation from whole-genome sequencing. *Mol. Ecol. Res.* 14, 199–208.
- Wagner-Wright, S. (1995). *History of the Macadamia Nut in Hawaii, 1881-1981: From Bush Nut to Gourmet's Delight*. Lewiston, ID: Edwin Mellen Press.
- Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., and Sugiura, M. (1994). Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. U.S.A.* 91, 9794–9798. doi: 10.1073/pnas.91.21.9794
- Weber, L. C., Vanderwal, J., Schmidt, S., McDonald, W. J., and Shoo, L. P. (2014). Patterns of rain forest plant endemism in subtropical Australia relate to stable mesic refugia and species dispersal limitations. *J. Biogeog.* 41, 222–238. doi: 10.1111/jbi.12219
- Whittall, J. B., Syring, J., Parks, M., Buenrostro, J., Dick, C., Liston, A., et al. (2010). Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Mol. Ecol.* 19, 100–114. doi: 10.1111/j.1365-294X.2009.04474.x
- Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., et al. (2005). The effects of artificial selection on the maize genome. *Science* 308, 1310–1314. doi: 10.1126/science.1107891
- Yamori, W., Makino, A., and Shikanai, T. (2016). A physiological role of cyclic electron transport around photosystem I in sustaining photosynthesis under fluctuating light in rice. *Sci. Rep.* 6:20147. doi: 10.1038/srep20147
- Zhang, H., Miao, H., Wang, L., Qu, L., Liu, H., Wang, Q., et al. (2013). Genome sequencing of the important oilseed crop *Sesamum indicum* L. *Genome Biol.* 14, 401. doi: 10.1186/gb-2013-14-1-401
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., and Song, B. H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evol. Appl.* 10, 5–24. doi: 10.1111/eva.12434
- Zhao, Y., Yin, J., Guo, H., Zhang, Y., Xiao, W., Sun, C., et al. (2015). The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. *Front. Plant Sci.* 5:696. doi: 10.3389/fpls.2014.00696

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Nock, Hardner, Montenegro, Ahmad Termizi, Hayashi, Playford, Edwards and Batley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.