

Article

Genetic Structure of Wild Germplasm of Macadamia: Species Assignment, Diversity and Phylogeographic Relationships

Thuy Mai *, Mobashwer Alam, Craig Hardner, Robert Henry and Bruce Topp

Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, QLD 4072, Australia; m.alam@uq.edu.au (M.A.); c.hardner@uq.edu.au (C.H.); robert.henry@uq.edu.au (R.H.); b.topp@uq.edu.au (B.T.)

* Correspondence: thi.mai3@uqconnect.edu.au

Received: 15 May 2020; Accepted: 30 May 2020; Published: 3 June 2020



Abstract: Macadamia is an Australian native rainforest tree that has been domesticated and traded internationally for its premium nuts. Common cultivars rely upon a limited gene pool that has exploited only two of the four species. Introducing a more diverse germplasm will broaden the genetic base for future crop improvement and better adaptation for changing environments. This study investigated the genetic structure of 302 accessions of wild germplasm using 2872 SNP and 8415 silicoDART markers. Structure analysis and principal coordinate analysis (PCoA) assigned the 302 accessions into four distinct groups: (i) *Macadamia integrifolia*, (ii) *M. tetraphylla*, and (iii) *M. janseni* and *M. ternifolia*, and (iv) admixtures or hybrids. Assignment of the four species matched well with previous characterisations, except for one *M. integrifolia* and four *M. tetraphylla* accessions. Using SNP markers, 94 previously unidentified accessions were assigned into the four distinct groups. Finally, 287 accessions were identified as pure examples of one of the four species and 15 as hybrids of *M. integrifolia* and *M. tetraphylla*. The admixed accessions showed the highest genetic diversity followed by *M. integrifolia*, while *M. ternifolia* and *M. janseni* accessions were the least diverse. Mantel test analysis showed a significant correlation between genetic and geographic distance for *M. integrifolia* ($r = 0.51$, $p = 0.05$) and a positive but not significant correlation for *M. tetraphylla* ($r = 0.45$, $p = 0.06$). This study provides a population genetics overview of macadamia germplasm as a background for a conservation strategy and provides directions for future macadamia breeding.

Keywords: genetic diversity; DART markers; macadamia; dendrogram; principal coordinate analysis; population structure; population genetics; wild species

1. Introduction

The genus *Macadamia* belongs to the Proteaceae family and is composed of four species: *Macadamia integrifolia* Maiden and Betche, *M. janseni* C.L. Gross and P.H. Weston, *M. ternifolia* F. Muell and *M. tetraphylla* L.A.S. Johnson. The natural distribution of the four species is in the subtropical rainforest from south-east Queensland (QLD) to north-east New South Wales (NSW), Australia [1,2]. *M. integrifolia* and *M. ternifolia* occur in south-east QLD, while *M. tetraphylla* is distributed mainly in northern NSW, with some extension into south-east QLD [3]. Overlapping zones exist between *M. integrifolia* and *M. tetraphylla* and between *M. integrifolia* and *M. ternifolia*, with natural hybridisation occurring in these zones [4]. The fourth species, *M. janseni*, has been found only in a single location in Bulburin (QLD) that is 180 km north from the nearest population of *M. integrifolia* [5]. Of these four species, *M. integrifolia* and *M. tetraphylla* produce edible nuts, and hence, most of the commercial cultivars belong to either of these two species or their hybrids [6]. The other two species, *M. janseni* and

M. ternifolia, have not been used in directed breeding, due to their bitter inedible nuts containing high levels of cyanogenic glycoside [7–9]. Most current commercial varieties appear to be two to four generations from the wild [1]. The majority of global macadamia production relies upon the cultivars from the Hawaiian breeding program, which is comprised mostly of a limited gene pool of *M. integrifolia* [10,11]. However, wild genetic resources have the potential to provide parents with desirable traits, including small tree size, nuts with thinner shells, resistance/ tolerance to biotic and abiotic factors, etc. [12] Exploring the genetic potential of the wild germplasm will facilitate further exploitation of these resources in trait improvement.

The study of genetic diversity and population structure can determine the potential of wild germplasm in future breeding. The genetic structure is formed over time due to the multiple actions of migration, selection, mutation, and genetic drift [13], as well as the mode and method of reproduction. A diverse species has the opportunity for selection of the fittest alleles while low diversity leads to the risk of extinction [14]. Broadening the genetic base of breeding material requires the identification of diverse parents for crossing with the cultivated crop [14]. Understanding the genetic relationship among the parents is essential to avoid inbreeding depression, particularly for the improvement of complex traits. Therefore, knowledge of the genetic divergence is a prerequisite to maximise heterosis in the breeding progeny [15]. The history of world macadamia breeding is very short, and the existing cultivars are only a few generations from the wild germplasm growing naturally in the rain forest. Recently, a chloroplast genome sequencing project on wild and cultivated germplasm indicated that all major Hawaiian cultivars share a single chlorotype probably derived from a small sample from single location [16]. As current macadamia plantations are mostly dependent on Hawaiian cultivars, this limited sampling suggests there is an opportunity for future genetic improvement by exploiting the diverse genetic resource. In addition, the genetic base can be expanded in breeding material by selecting diverse parents from wild germplasm for crossing with the cultivated crop.

Molecular markers are considered as the most suitable tool to estimate genetic diversity, due to their polymorphic nature and independence to environmental effects [17]. Several molecular marker systems have been developed to study in macadamia and only a few of them were used for the genetic characterisation of wild germplasm. An isozyme-based study was conducted by Aradhya et al. [18] to identify the genetic relationships among 40 cultivars (35 *M. integrifolia*, three *M. tetraphylla*, and two hybrids) and three *M. ternifolia* accessions. Mast et al. [19] studied the relationships among the four *Macadamia* species and their wild relatives, using three cpDNAs (*matK*, *atpB*, and *ndhF*), three nDNA (*waxy* loci 1 and 2, and *PHYA*) genomic regions. However, this study used only one accession per species. The genetic structure of a large number of wild germplasm accessions was studied by Peace [4] using low throughput RAF (randomly amplified DNA fingerprinting, dominant) and RAMiFi (randomly amplified microsatellite fingerprinting, co-dominant) markers. All these marker systems have limitations, including a low total number of markers, low marker density, and low genome coverage, and hence, are seldom used in genomic studies. SSRs (simple sequence repeat markers or microsatellite) are considered as one of the best marker systems for genetic studies, with many advantages, such as stability, PCR-based amplification, and relatively low cost [17]. SSRs were used in the genetic diversity study of wild *M. integrifolia* [20] and wild *M. tetraphylla* populations [21,22]. However, there is a limited number of SSR primers available for macadamia, particularly those that successfully amplify across species [23], and as such, may not be effective for a large-scale genetic study of all four wild species.

The rapid advancement of next generation sequencing technology (NGS) enables the discovery of high-throughput and cost-effective molecular marker systems. Using NGS technology, Diversity Array Technology (DArT) developed a marker system that facilitates affordable whole-genome level genetic characterisation. DArT has been successfully applied for the genetic diversity, population structure and genetic mapping studies of many crop species [24–26]. Recently, Alam et al. [27] used 11,526 silicoDArTs and 3956 SNPs to study the genetic diversity and population structure of 80 macadamia cultivars. O'Connor et al. [28] reported the genetic diversity, population structure and linkage disequilibrium

of 295 seedling progenies from 29 selected parents, using 16,171 silicoDArTs and 4113 SNPs. These studies suggest that DArTseq markers could be applied for genomic studies in the wild germplasm of macadamia.

In this study, for the first time, we used high-throughput DArTseq platforms for the genetic characterisation of a large number of wild accessions of macadamia. The aims were to: (1) assess the population structure of wild macadamia germplasm, (2) explore the genetic diversity among the accessions within species, and (3) determine the relationship between genetic and geographic distance within *M. integrifolia* and *M. tetraphylla*.

2. Results

2.1. Quality of DArTseq Markers

DArTseq platforms generated 13,221 SNP and 47,811 silicoDArT markers. The call rates of SNP markers varied from 0.20 to 1.00, with an average of 0.62 (Supplementary Materials Table S1). Of the 13,221 SNPs, the call rate of 4184 markers (32%) was > 0.80 (Figure 1a). The reproducibility of SNPs varied from 0.86 to 1.00, and most of them (98%) were over 0.95 (Figure 1a). The call rate of silicoDArT markers varied from 0.81 to 1.00 (Table S1). The average call rate was very high (>0.99), with 94% of markers having a call rate over 0.95 (Figure 1a). The range of reproducibility was 0.95 to 1.00, in which 98% of the markers had very high value (0.99) of reproducibility. Mean one ratio was higher in SNPs (0.32) than in silicoDArT markers (0.08) (Table S1). Most silicoDArTs (82%) had one ratio below 0.05, compared with only 28% of SNPs (Figure 1b). Considering the quality parameters: call rate (>0.80), reproducibility (>0.95) and one ratio (>0.05), 2872 SNPs and 8415 silicoDArTs were retained for further analysis (Tables S2 and S3). The remaining markers had PIC values from 0 to 0.5 for both SNPs and silicoDArTs (Figure 1c). Mean PIC was 0.26 for silicoDArTs and 0.22 for SNPs. Only 120 of the SNPs (4%) had low PIC (<0.05), compared with 825 (9.8%) of silicoDArTs (Figure 1c).

2.2. Population Assignment

The K and Q values from STRUCTURE analysis were used for the assignment of individual accessions in each species/hybrid group. Population clusters in PCoA validated the species representation.

The ΔK from the STRUCTURE analysis of SNP markers was significant when $K = 2, 3$ and 5 , with a peak at $K = 3$ (Figure 2a). The optimal peak at $K = 3$ suggested that the 302 accessions in the germplasm were derived from three distinct clusters, as represented by different colours in the structure analysis (Figure 2b, $K = 3$), here named Cluster I (blue), Cluster II (green) and Cluster III (red). Cluster I was composed of 18 accessions, including eight previously labelled as *M. ternifolia*, two as *M. janseni*, and eight as undefined species (Table S4). These undefined eight accessions, originally labelled as mixed/hybrid populations, were collected from the natural distribution of *M. ternifolia*. Considering this distribution and their morphological appearance (Thuy Mai, pers. observations), these accessions were classified as *M. ternifolia*. Cluster II contained 99 predefined *M. integrifolia* and 36 accessions of previously undefined species. Cluster III was comprised of 94 predefined *M. tetraphylla* and 38 undefined species. There were 17 accessions, including one predefined *M. integrifolia*, four predefined *M. tetraphylla*, and 12 accessions from planted/unknown/mixed populations that showed the genetic admixture (hybrid) of clusters. For example, the accession "M034", which was previously labelled as *M. integrifolia*, consisted of 6% of Cluster II (predominantly *M. integrifolia*), 16% of Cluster III (predominantly *M. tetraphylla*), but 78% of Cluster I (predominantly *M. janseni* and *M. ternifolia*). Two previously labelled *M. tetraphylla* accessions ("M265" and "M266") were identified as admixtures of Cluster II & III. The accession "M265" was composed of 61% of Cluster II and 39% of Cluster III, and accession "M266" was composed of 52% of Cluster III.

The pattern displayed for $K = 2$ (Figure 2b) grouped the accessions of three species, *M. janseni*, *M. ternifolia* and *M. tetraphylla*, into one cluster, and separated *M. integrifolia* accessions in another

cluster. The pattern displayed for $K = 5$ (Figure 2b) still did not separate the accessions of the two species *M. jansanii* and *M. ternifolia* but divided the accessions of *M. integrifolia* into smaller sub-clusters.

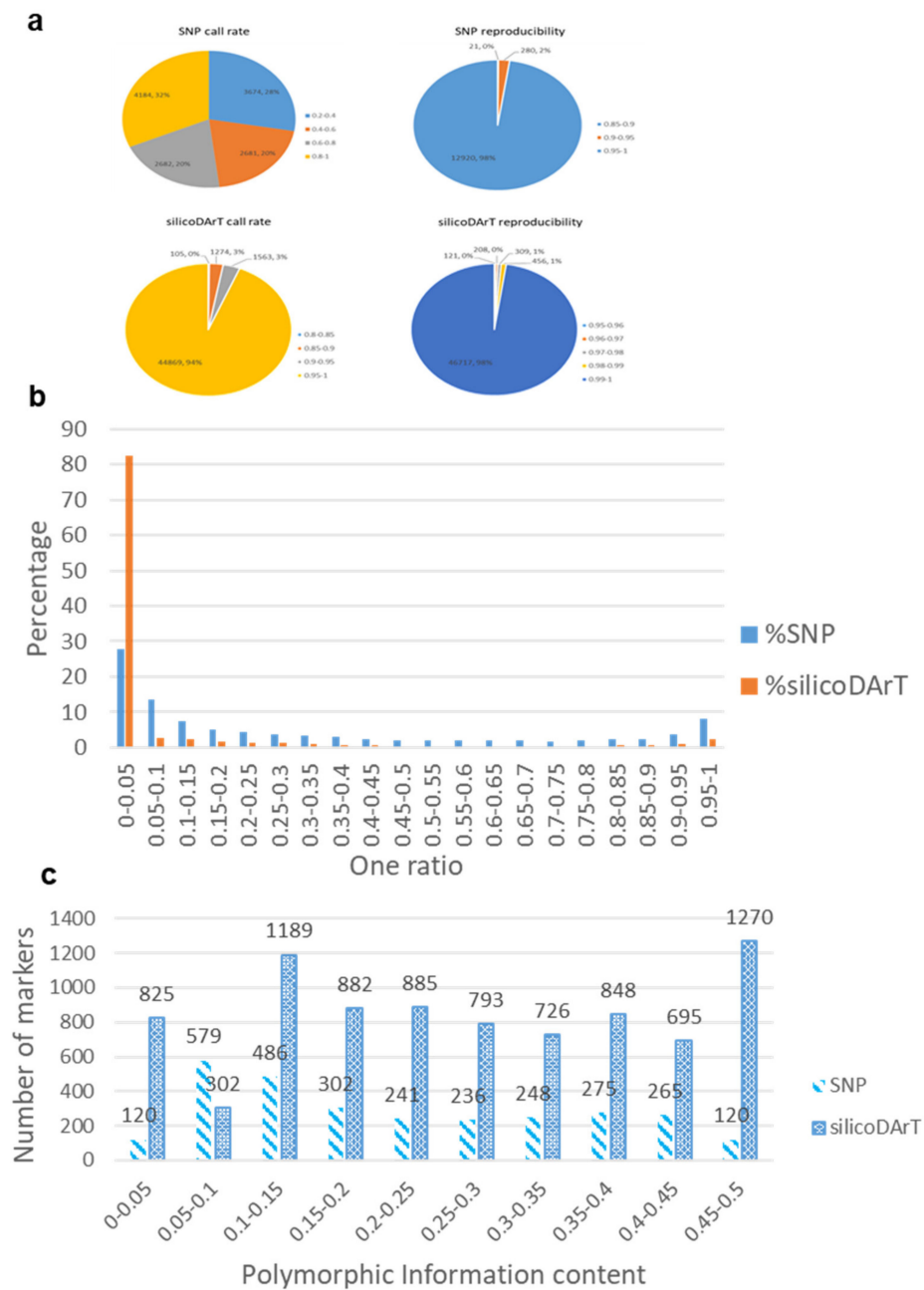


Figure 1. Distribution of SNP and silicoDArT marker data for several quality parameters. (a) call rate and reproducibility; (b) one ratio and (c) polymorphic information content (PIC) value.

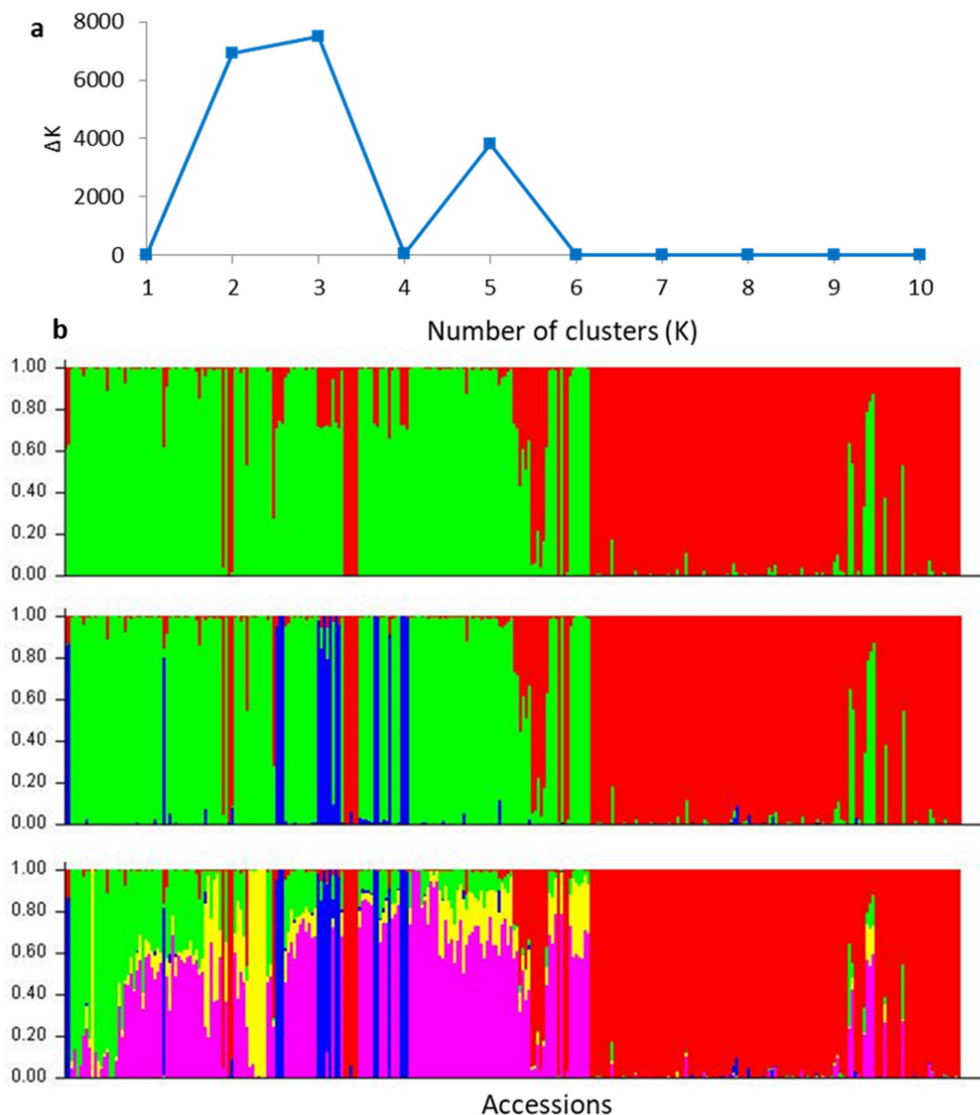


Figure 2. Population structure of 302 accessions based on 2872 SNPs, as inferred by STRUCTURE. (a) Best value of K based on Evanno's ΔK ; (b) Individual membership proportions (Q value) in two, three and five clusters, with each cluster represented by a colour block. Each vertical line represents one accession. The accessions were sorted on the x-axis by latitude from north to south.

Principal coordinate analyses (PCoA) of SNP markers via distance matrix with data standardization identified three distinct groups of the four species (Figure 3). This result was consistent with the result of STRUCTURE analysis at $K = 3$. The first two coordinates of PCoA explained 61.09% of total variation in SNPs. The accessions of two species *M. ternifolia* and *M. jansenii* formed Cluster I. Cluster II was formed by the accessions of *M. integrifolia* and Cluster III included the accessions of *M. tetraphylla*. The accession "M034", which was assigned as admixture in STRUCTURE analysis, was clustered in the *M. jansenii*/*M. ternifolia* group and shows a close relationship with two *M. jansenii* accessions. The accession "M160", which was also assigned as admixture in STRUCTURE analysis composing 78% of Cluster III, was clustered in the *M. tetraphylla* group.

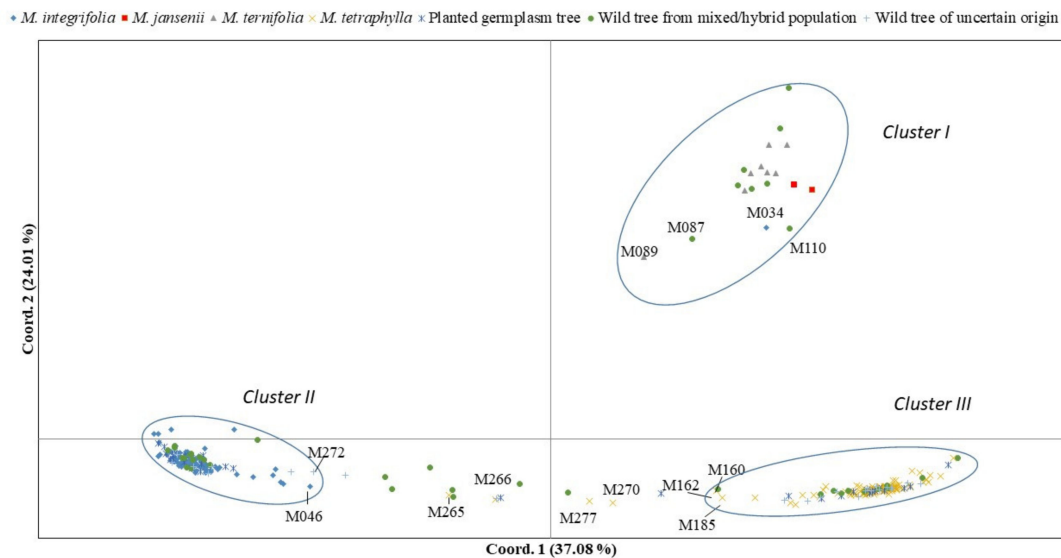


Figure 3. Principal coordinate analysis (PCoA) of the 302 accessions based on 2872 SNP markers, showing the three distinct groups and the admixtures. The first two coordinates of PCoA explained 34.06% of the total variation.

Finally, based on the STRUCTURE analysis and PCoA, we assigned 302 wild accessions into 287 pure accessions, representing the four distinct species, and 15 admixtures. Pure accessions are composed of 135 *M. integrifolia*, 133 *M. tetraphylla*, and 19 *M. ternifolia*/*M. jansenii* (Table S4).

2.3. Genetic Diversity

We estimated the genetic diversity parameters among 302 accessions, using both SNP and silicoDArT markers (Table 1). For SNP markers, the number of effective alleles (N_e) ranged from 1.08 to 1.34, with the lowest in *M. ternifolia*/*M. jansenii* group and the highest in admixture. Shannon's index (I) ranged from 0.11 (*M. ternifolia*/*M. jansenii*) to 0.33 (admixture), with a mean of 0.23. In all clusters, the observed heterozygosity (H_o) was smaller than the expected heterozygosity (H_e). H_e was highest in admixture (0.21) and lowest in *M. ternifolia*/*M. jansenii* (0.07), with a mean of 0.15. Interestingly, accessions from the *M. integrifolia* group showed the highest percentage (86.53%) of polymorphism (%P), followed by admixture (74.93%), *M. tetraphylla* (71.5%), *M. ternifolia*/*M. jansenii* (32%) groups. Similar results were also observed for silicoDArT markers (Table 1).

Nei's genetic distance (D), based on SNP markers (Table 2), ranged from 0.06 between admixtures and *M. integrifolia* to 0.27 between *M. integrifolia* and *M. ternifolia*/*M. jansenii* groups. *M. tetraphylla* accessions shows lower genetic distance with *M. integrifolia* ($D = 0.2$) than that of *M. ternifolia*/*M. jansenii* accessions ($D = 0.23$). Hence, the admixture accessions showed almost similar genetic distance with *M. integrifolia* (0.06) and *M. tetraphylla* (0.07) germplasm. Although the estimated value of genetic distance using silicoDArT markers was lower than that of SNPs, the genetic relationship between species was similar in both marker systems (Table 2).

The analysis of molecular variance (AMOVA) showed a higher proportion of variance detected within species than among clusters (Table 3). For SNPs, the percentage of genetic variation within species (55%) was higher than that among species (45%). A similar pattern of genetic variation was observed using silicoDArT markers (Table 3).

Table 1. Genetic diversity parameters for Macadamia accessions based on SNP and silicoDArT markers. N = number of accessions, Na = number of different alleles, Ne = number of effective alleles, I = Shannon's information index, Ho = observed heterozygosity, He = expected heterozygosity, %P = percentage of polymorphic loci, SE = standard error.

Group		N	Na	Ne	Ho	He	I	%P
SNPs								
<i>M. ternifolia</i> / <i>M. janseni</i> (Cluster I)	Mean	19	1.28	1.08	0.04	0.07	0.11	32.00
	SE		0.01	0.01	0.00	0.00	0.00	
<i>M. integrifolia</i> (Cluster II)	Mean	135	1.87	1.28	0.12	0.18	0.29	86.53
	SE		0.01	0.01	0.00	0.00	0.00	
<i>M. tetraphylla</i> (Cluster III)	Mean	133	1.72	1.20	0.08	0.12	0.20	72.32
	SE		0.01	0.01	0.00	0.00	0.00	
Admixture	Mean	15	1.75	1.34	0.18	0.21	0.33	74.93
	SE		0.01	0.01	0.00	0.00	0.00	
Total	Mean	302	1.66	1.22	0.11	0.15	0.23	47.20
	SE		0.01	0.00	0.00	0.00	0.00	19.30
silicoDArTs								
<i>M. ternifolia</i> / <i>M. janseni</i> (Cluster I)	Mean	19	0.58	1.11	-	0.07	0.11	24.01
	SE		0.01	0.00	-	0.00	0.00	
<i>M. integrifolia</i> (Cluster II)	Mean	135	1.66	1.35	-	0.21	0.33	79.94
	SE		0.01	0.00	-	0.00	0.00	
<i>M. tetraphylla</i> (Cluster III)	Mean	133	1.47	1.26	-	0.16	0.24	70.86
	SE		0.01	0.00	-	0.00	0.00	
Admixture	Mean	15	1.53	1.36	-	0.22	0.34	70.18
	SE		0.01	0.00	-	0.00	0.00	
Total	Mean	302	1.31	1.27	-	0.16	0.26	61.25
	SE		0.01	0.00	-	0.00	0.00	12.61

Table 2. Pairwise population matrix of Nei's genetic distance (D) among the clusters of Macadamia wild germplasm, using SNP and silicoDArT markers.

Group	<i>M. integrifolia</i>	<i>M. tetraphylla</i>	<i>M. ternifolia</i> / <i>M. janseni</i>
SNPs			
<i>M. tetraphylla</i>	0.20		
<i>M. ternifolia</i> / <i>M. janseni</i>	0.27	0.23	
Admixture	0.06	0.07	0.22
silicoDArTs			
<i>M. tetraphylla</i>	0.17		
<i>M. ternifolia</i> / <i>M. janseni</i>	0.17	0.16	
Admixture	0.08	0.05	0.16

Table 3. Summary statistic of AMOVA analysis in Macadamia germplasm using SNP and silicoDArT markers. df = degrees of freedom, SS = sum of squared observations, MS = mean of squared observations, Est. Var = estimated variance, % Var. = percentage of total variance. PhiPT = var. among groups (species) / (var. among groups + var. within group).

Source	df	SS	MS	Est. Var.	% Var.	PhiPT Statistic	p-Value
SNPs							
Among species	3	175,814	58,604	951	45%		
Within species	298	344,429	1155	1155	55%		
Total	301	520,243		2107	100%	0.45	0.001
silicoDArTs							
Among species	3	74,996	24,999	401	34%		
Within species	298	235,037	789	789	66%		
Total	301	310,033		1190	100%	0.34	0.001

2.4. Phylogeographic Relationships among the Accessions of *M. integrifolia* and *M. tetraphylla*

We identified the genetic relationships within each species of *M. integrifolia* and *M. tetraphylla* presented in the dendrogram (Figure 4). Most accessions from the same locality grouped together, although some accessions from a locality clustered with accessions from other localities. For example, within *M. integrifolia*, the accession “M010” from Gundiah and four accessions from Nambour grouped with Beenleigh accessions (Figure 4a). Two accessions, “M027” and “M037”, from Gympie grouped with accessions from Gundiah and Nambour, respectively. Within *M. tetraphylla*, the accession “M263” from Lismore grouped with Murwillumbah, while the accessions “M208” and “M209” from Murwillumbah grouped with Lismore (Figure 4b). Some of the accessions showed an unexpectedly variable branch length compared to other accessions from the same cluster. Accessions with the longest branches represent the most diverged accessions within the population. For example, accession “M148” had ~50% longer branch than other Beenleigh accessions. Similarly, accessions “M036”, “M048”, “M136” and “M008” of *M. integrifolia*, and “M227”, and “M267” of *M. tetraphylla* had significantly longer branches.

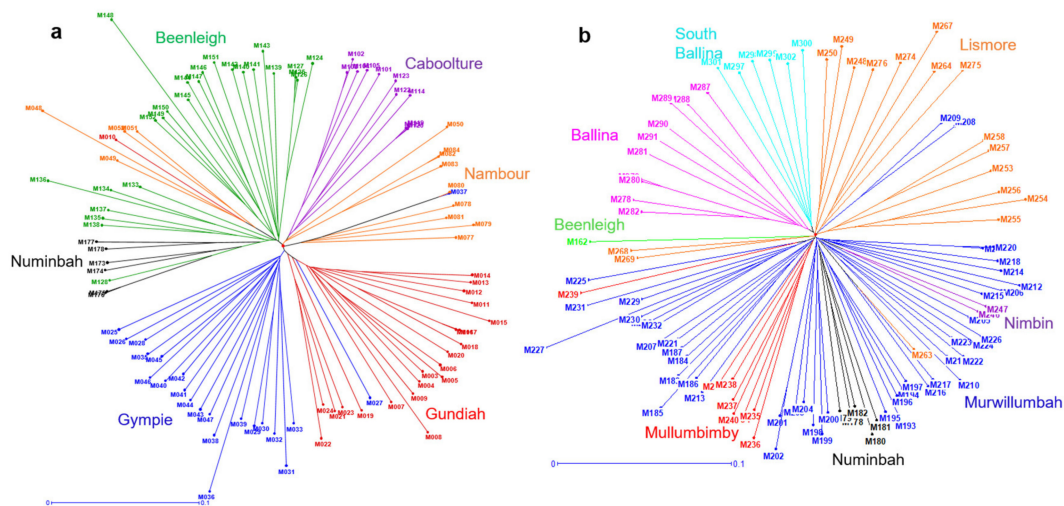


Figure 4. Unweighted neighbour-joining dendrograms using 2872 SNP markers, showing the genetic relationships among (a) 99 *M. integrifolia* accessions and (b) 94 *M. tetraphylla* accessions. Localities are represented by different colours.

To explore the genetic basis of the geographic relationship we calculated the correlation between the genetic distance and geographic distances of the localities for each species. The pairwise genetic distance among the accessions of six localities of *M. integrifolia* (Table S5) varied from 0.019 to 0.055, with an average of 0.034. The genetic distance between Gundiah and Gympie was the closest (0.019), while it was the farthest (0.055) between the accessions of Numinbah and Caboolture. The pairwise genetic distances among the accessions of *M. tetraphylla* localities (Table S6) were higher than that of *M. integrifolia*. The range of variation in *M. tetraphylla* was 0.013 to 0.151, with a mean of 0.063. Accessions from locality Murwillumbah and Lismore showed the closest genetic distance (0.013), while the highest (0.151) was observed between the accessions of Beenleigh and Nimbin.

The Mantel test analysis showed a significant correlation between genetic distance and geographic distance ($r = 0.51$, $p = 0.05$) among *M. integrifolia* localities (Figure 5a). Meanwhile, the correlation between genetic and geographic distance among *M. tetraphylla* localities was positive but not significant ($r = 0.45$, $p = 0.06$) (Figure 5b).

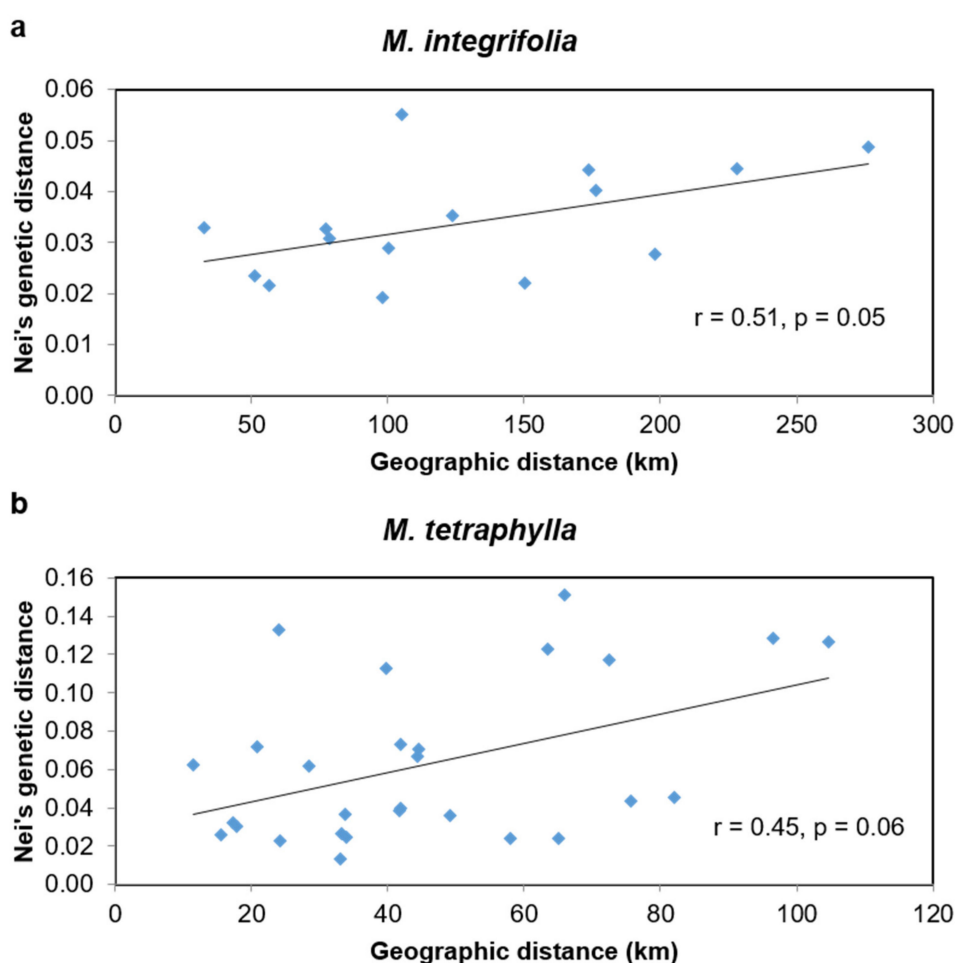


Figure 5. Correlation between genetic and geographical distance among (a) six localities of *M. integrifolia*, and (b) eight localities of *M. tetraphylla*, based on a Mantel test at 999 random permutations.

3. Discussion

3.1. Species Assignment of Wild Macadamia Germplasm

Population structure and PCoA of DArTseq based SNP markers facilitated the assignment of individuals in corresponding species or hybrid groups. Our species classification of each accession matches well with the field note and previous DNA study [4]. Results from this study confirmed that most of the previous phenotypic characterisations were successful for species' identification of wild germplasm.

All but one *M. integrifolia* accessions were clustered together in the same group. "M034", an accession from the "Mooloo" region of the "Gympie" locality, was identified as *M. ternifolia*/*M. jansanii*, although it was previously recorded as *M. integrifolia*. A co-investigation of wild germplasm using 15 SSR markers identified that the same accession "M034" is a clone of another accession (X-CANB896104) from Canberra Botanic Garden. Interestingly, accession "X-CANB896104" is recorded as a cutting from wild *M. ternifolia* from Mary Cairncross Park, Maleny (Cathy Nock., pers communication). Further phenotypic characterisation confirmed that "M034" is a *M. ternifolia*. Among the 99 accessions of *M. tetraphylla*, four accessions were assigned as hybrid of *M. tetraphylla* and *M. integrifolia*. Accessions "M265" and "M266" shared almost 50% from each species, whereas "M270" and "M277" had ~30% from *M. integrifolia* genotype and ~70% *M. tetraphylla* (Figure 3, Table S3).

This study clearly demonstrated that the accessions of *M. tetraphylla* and *M. integrifolia* formed two distinct populations. The accessions of *M. jansanii* and *M. ternifolia* clustered together, although

they were collected from geographically distant locations. The close relationship between *M. jansanii* and *M. ternifolia* is supported by the previous molecular studies [29,30]. These findings indicate that the accessions of both *M. jansanii* and *M. ternifolia* may have the same genetic lineage, or that one may be an ancestor of the other. Possibly, these two small groups of populations may have separated due to past climatic extremes and adapted as small groups in two distinct locations. It is to be noted that there may be some sampling effect of the small number of accessions of *M. jansanii* ($n = 2$) and *M. ternifolia* ($n = 16$) in our study. Though *M. jansanii* and *M. ternifolia* differ in a few simple traits including leaf tip, leaf serration and flower colour, they are morphologically similar, both with small tree size, and small and bitter nuts [9,31]. However, there may be further debate on the species differentiation of these two small populations. Investigation on evolutionary genetics and time divergence on four *Macadamia* species can be used to confirm the speciation of wild macadamia germplasm.

The species status of previously unidentified accessions, including those of unknown origins, planted germplasm and mixed/hybrid populations were resolved. Out of 94 unidentified accessions, 36 were assigned as *M. integrifolia*, 39 as *M. tetraphylla*, 8 as *M. ternifolia* and 11 as hybrids/admixtures of *M. integrifolia* and *M. tetraphylla* (Table S5). The genotypic classification was consistent with our field observation on phenotypic characteristics of representative species (Thuy Mai, pers communication). The species composition of many modern and heritage cultivars is uncertain [1], but this finding supports the potential of SNP markers to resolve their species status.

Some accessions we identified as *M. tetraphylla* had been collected from further north than the accepted distribution of this species. These were the accessions “M056”, and “M057” from population 16 (Palmwoods, Nambour QLD), accession “M054” from population 106 (Mapleton Kenilworth, Nambour QLD) and five accessions from population 36 (Mount Glorious, Caboolture QLD). However, these populations were noted at the time of collection as planted or uncertain populations, and it seems highly likely that their locations were the result of human activity.

3.2. Genetic Diversity of the Four *Macadamia* Species

In this study, we developed new knowledge on genetic diversity in wild accessions by using high-throughput silicoDArT and SNPs marker. In both types of markers, the average expected heterozygosity (H_e) across the wild germplasm and clusters forming “pure” species was lower than in previous genetic studies of macadamia cultivars [21–23,32]. However, the cluster of hybrid accessions from wild *M. integrifolia* and *M. tetraphylla*, which is composed of a small number of accessions ($n = 15$), showed greater gene diversity than previous studies on cultivars and pure wild species in the current study. This result indicates that crossing between two species can be conducted, to increase genetic diversity in the future breeding program.

Our results suggested that *M. integrifolia* and *M. tetraphylla* contained two- to three-fold greater diversity than the *M. jansanii* and *M. ternifolia* cluster. Population size has a significant effect on genetic diversity [33]. Generally, smaller population size leads to lower genetic diversity [34]. Extinction and contraction of species’ distribution during successive ice ages has resulted in reduced population size and resultant diversity bottlenecks in other Australian flora, such as *Acacia*, *Banksia*, *Eucalyptus* etc. [35]. Certainly, *M. jansanii* formed a very small population, with less than 100 individuals comprising the whole species [5,9]. In this study, only a small number of *M. jansanii* ($n = 2$) and *M. ternifolia* ($n = 17$) have been tested. The lower diversity of *M. jansanii* and *M. ternifolia* in our study may be the result of small populations. An investigation with larger sample numbers from diverse distributions of *M. jansanii* and *M. ternifolia* should be conducted, to define their genetic diversity more completely and accurately.

3.3. Phylogeographic Relationship among the Accessions of *M. integrifolia* and *M. tetraphylla*

Geographical distance is one of the major contributing factors in species differentiation. Knowledge of the genetic structure of a species over its geographic distribution is important to develop an understanding of the evolutionary processes [36]. In this study, the neighbour-joining tree, based on

dissimilarity matrix (Figure 6), showed that the accessions from the same geographical area appeared to be grouped together. However, some accessions (e.g., “M010”, “M027”, “M037”, “M263”, “M208”, “M209”) were found to be clustered with accessions of different geographical regions (Figure 6). This result was supported by a previous study on chloroplast genome sequence, where Nock et al. [16] reported the relocation of some accessions within the *M. integrifolia* germplasm. Since the gene flow for both species is restricted within a short distance (~50 km) [1], the impact of environmental parameters, such as water, gravity, and animals like rodents [37], or a result of human activity on the seed transportation [16], could be considered as the cause of relocation.



Figure 6. Map showing origins of representative wild-germplasm accessions of *M. integrifolia* and *M. tetraphylla*. The black circles represent six localities of *M. integrifolia* population, and the red circles represent eight localities of *M. tetraphylla* populations.

Phylogeographic analysis of *M. integrifolia* accessions revealed a significant positive correlation between genetic and geographic distance ($r = 0.51, p = 0.05$). This correlation was higher than a previous study using RAF markers [4] ($r = 0.16, p = 0.016$), and in almond germplasm ($r = 0.173, p = 0.226$) [38]. Based on chloroplast genome study, Nock et al. [16] also found a latitudinal population structure among these accessions. We also observed a positive relationship between genetic and geographic distance among the accessions of wild *M. tetraphylla*. Although, this correlation ($r = 0.45$) is non-significant, it is higher than that found by Peace [4] ($r = 0.13$). The non-significant phylogeographic relationship in *M. tetraphylla* suggested that geographic distance may not be the main factor influencing the genetic distance between populations of this species, although geographical boundaries, low gene flow and genetic drift are typically key factors explaining genetic differentiation in fragmented populations [39].

The phylogenetic trees (Figure 4) of *M. integrifolia* and *M. tetraphylla* provided an indication of ancestral lineage of the accessions of each species. Interestingly, the root of most of the accessions of both species originated from the population of Numinbah (Figure 4). It is to be noted that Numinbah is an overlapping region of both species, and its surrounding regions are the sources of mixed/hybrid population. We hypothesise that there is a possibility of early divergence of these two species at Numinbah. Later, smooth leaved *M. integrifolia* may have adapted to the north and serrated leaved *M. tetraphylla* may have adapted to the southern regions. Further genomic investigation with more accessions from Numinbah can explain the origin of these two species.

4. Materials and Methods

4.1. Collection of Plant Samples

The germplasm field trials were established in 2000 and 2001 in Tiaro, Queensland (QLD) and Alstonville, New South Wales (NSW), Australia. Wild accessions were collected from multiple geographical regions (Figure 6), covering the natural distribution of the four species. Ramets of each accession were propagated clonally as rooted cuttings. During the collecting trip, the germplasm collector (S. Faulkner) classified the populations as “wild”, “planted” (trees were cultivated from the local wild trees), “hybrid/mixed”, or “uncertain”. The accession(s) from each population may belong to more than one of these classifications; therefore, the original description for each accession was noted. Accessions of the rare species *M. janseni* were added to the field trial in July 2011. From these collections, we studied 302 accessions, including 100 *M. integrifolia*, two *M. janseni*, eight *M. ternifolia*, 98 *M. tetraphylla*, and 94 accessions of undefined species (from mixed/hybrid populations, uncertain origins or planted germplasm). These accessions originated from 75 populations (one to seven accessions per population, averaging 4.1) across 52 regions (one to three populations per region) from 14 localities (one to nine regions per locality) (Table S7). The accessions were ordered by latitude from north to south and coded from “M001” to “M302”.

4.2. DNA Extraction and Genotyping

Newly flushed young leaves were collected from the ex situ trials at Tiaro and Alstonville in December 2017 and placed in zipped plastic bags inside a cool box with ice blocks. The materials were stored in a cold room at 4 °C before transfer to DArT Pty Ltd. (Canberra, Australian Capital Territory, Australia), to perform DNA extraction. A total of 2–3 mg of leaf tissue from each sample was sub-sampled into a 1.1 mL microtube containing a disposable steel ball bearing. Leaf samples were crushed using a QIAGEN Tissue Lyser (Qiagen, Germany). A Freedom EVO robotic Tecan 100 (Tecan Group, Switzerland) was used for DNA extraction following an existing protocol recommended by DArT (<https://www.diversityarrays.com/>). DNA samples were incubated with loading dye at 37 °C for two hours and then checked for quality control on 0.8% agarose electrophoresis gel for 30 min at 100 V.

Accessions were genotyped for dominant silicoDArT and co-dominant SNP markers following an established protocol developed by Kilian et al. [40] An appropriate method (*Pst*I + *Hha*I) of complexity reduction was selected to detect DArTseq-based markers using next generation sequencing (NGS) technology. The silicoDArT and SNP markers were scored as a binary data format in which the

score was “1” for presence, “0” for absence and “-”, for failure to score of a marker in the genomic representation of each sample. The full details of methodology of developing DArTseq-based markers for macadamia were previously described by Alam et al. [27] and O’Connor et al. [28]

4.3. Analysis of Processed Marker Data

DArTseq platforms generated SNP and silicoDArT markers. DArTsoft v7.4 software was used to automatically identify and score the polymorphic markers, using a proprietary marker calling algorithms. The quality of the markers was tested for call rate (%), reproducibility (%), one ratio, and polymorphic information content (PIC). The call rate determines the success of reading the marker sequence across the samples, and was estimated from the percentage of samples for which the score was either “0” (absence of marker) or “1” (presence of marker). The scoring of reproducibility involves the proportion of 138 technical replicates for which the marker score exhibited consistency. One ratio was determined as the proportion of samples for which genotypes were scored as “1”. PIC is the degree of diversity of the marker in the population and shows the usefulness of the marker for linkage analysis. Quality control and filtering were applied to both SNP and silicoDArT markers, including call rate (>80%), reproducibility (>95%), and one ratio (>0.05).

4.4. Analysis of Population Structure and Genetic Diversity

The population structure of the 302 accessions of four species was identified based on SNP markers by using the Bayesian model-based program STRUCTURE v2.3.4 [41–44]. A burn-in length of 10,000 cycles and Markov chain Monte Carlo (MCMC) of 50,000 runs were set for the structure analysis. Cluster values (K) ranging from one to 10 were performed, each with ten different iterations. Results from STRUCTURE were uploaded to Structure Harvester [45], through <http://taylor0.biology.ucla.edu/structureHarvester/>, to determine the optimum number of populations (best value of K) using ΔK value, as described by Evanno et al. [46] The individual ancestry proportion (Q value) at the best K value was determined for each accession. Based on Q values, the accessions were identified as pure or admixed species: accessions with Q value more than 0.8 were considered as “pure” and accessions with Q value less than 0.8 were assigned as “admixture” [14,47].

The genetic diversity and principal coordinate analysis (PCoA) were performed using GenAlEx v6.5.2 software [48]. Both SNP and silicoDArT markers were used to calculate the diversity parameters, including the mean number of alleles per locus (N_a), number of effective alleles (N_e), observed and expected heterozygosity (H_o and H_e), Shannon’s information index of diversity (I) and the percentage of polymorphic loci (%P). Pairwise population matrix of Nei’s genetic distance [49] and the analysis of molecular variance (AMOVA) were also estimated.

4.5. Phylogeographic Relationships Analysis

M. integrifolia and *M. tetraphylla* are the two major species used in current breeding programs. In this study, we have identified a large number of accessions of *M. integrifolia* and *M. tetraphylla*. Therefore, a detailed study was undertaken on the genetic relationships and geographical diversity of those species. Only accessions identified as “pure” representatives of each species that were confirmed by structure analysis, including 99 *M. integrifolia* accessions from six localities (Figure 6, black circles), and the 94 *M. tetraphylla* accessions from eight localities (Figure 6, red circles), were used. The “pure” accessions collected from planted wild germplasm or unknown origin were not included. DARwin v6.0 software [50] was used to estimate pairwise Jaccard’s genetic dissimilarity indices using 2872 SNP markers. A dendrogram was constructed by clustering accessions, based on a dissimilarity matrix using the unweighted neighbour-joining method. Clade strength in the dendrogram was tested using 100 bootstraps.

The correlation between genetic distance and geographic distance was estimated. The Nei’s genetic distance matrix and the pairwise geographic distance among the localities were calculated using GenAlEx v6.5.2 [48]. The Mantel test was used to determine the correlation between genetic and geographic distance, using 999 random permutations.

5. Conclusions

This is the first study to investigate the genetic structure of a large collection of wild macadamia germplasm using thousands of high-throughput molecular markers. A total of 302 wild accessions were characterised, using 2872 SNP and 8415 silicoDArT markers. Our population structure and principal co-ordinate analyses identified three distinct populations, in which *M. janseni* and *M. ternifolia* formed a single cluster. The Nei's genetic distance analysis clearly demonstrated that *M. janseni* and *M. ternifolia* are related and showed greater heterozygosity in *M. ternifolia* than in *M. janseni*. However, the limited number of accessions available in this study from *M. janseni* and *M. ternifolia* limits the strength of our conclusion on the diversity and population structure of these species. We suggest that further analysis with more accessions from these two species should be conducted, to increase understanding of genetic diversity and clarify their classification as distinct species. We observed the significant correlation between genetic and geographic distance among *M. integrifolia* populations. Additionally, we were able to confirm the species identity of unknown wild accessions and suggested the use of these markers to resolve the unclear species composition of domesticated cultivars.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/9/6/714/s1>, Table S1. Summary of markers; Table S2. List of SNP markers; Table S3. List of silicoDArT markers; Table S4. Q value at K = 3, Table S5. Genetic distance of *M. integrifolia*; Table S6. Genetic distance of *M. tetraphylla*; Table S7. 302 accessions information.

Author Contributions: Conceptualization, B.T. and T.M.; Methodology, T.M. and M.A.; Software, T.M. and M.A.; Validation, T.M.; Formal Analysis, T.M.; Resources, T.M.; Data Curation, T.M. and M.A.; Writing—Original Draft Preparation, T.M.; Writing—Review and Editing, M.A., C.H., R.H., B.T. and T.M.; Visualization, T.M.; Supervision, B.T., M.A., R.H. and C.H.; Project Administration, B.T. and M.A.; Funding Acquisition, B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research 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. PhD scholarship for TM was provided by Vietnam International Education Development (The Key Biotechnology Program—Ministry of Agriculture and Rural Development) and The University of Queensland.

Acknowledgments: The authors acknowledge Julia Playford for initiating the germplasm collection. Thanks to Agnelo Furtado (QAAFI), Jodi Neal (DAF), Katie O'Connor (QAAFI), Cathy Nock (Southern Cross University), Diversity Arrays Technology staff, Macadamia team and DAF staff, for their guidance and comments.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: The SNP and silicoDArT markers generated and analysed during this study are obtainable from The University of Queensland's Institutional Data Access/Ethics Committee, but restrictions apply to the availability of these data. The dataset "DArTseq markers of wild macadamia species" is available at <https://doi.org/10.14264/uql.2018.395>, for researchers who meet the criteria for access to confidential data. Contact data@library.uq.edu.au.

References

1. Hardner, C.M.; Peace, C.; Lowe, A.J.; Neal, J.; Pisanu, P.; Powell, M.; Schmidt, A.; Spain, C.; Williams, K. Genetic resources and domestication of macadamia. In *Horticultural Reviews*; Janick, J., Ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2009; Volume 35, pp. 1–125.
2. Topp, B.L.; Nock, C.J.; Hardner, C.M.; Alam, M.; O'Connor, K.M. Macadamia (*Macadamia* spp.) Breeding. In *Advances in Plant Breeding Strategies: Nut and Beverage Crops*; Al-Khayri, J., Jain, S.D.J., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 221–251.
3. Powell, M.; Accad, A.; Shapcott, A. 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.* **2014**, *62*, 322–334. [[CrossRef](#)]
4. Peace, C.P. Genetic Characterisation of Macadamia with DNA Markers. Ph.D. Thesis, University of Queensland, Brisbane, Australia, 2004.
5. Shapcott, A.; Powell, M. Demographic structure, genetic diversity and habitat distribution of the endangered, Australian rainforest tree help facilitate an introduction program. *Aust. J. Bot.* **2011**, *59*, 215–225. [[CrossRef](#)]

6. Hardner, C.; Topp, B. *Macadamia integrifolia* and *M. tetraphylla*. In *Trees for Life in Oceania: Conservation and Utilisation of Genetic Diversity*; Lee, T., John, D., Clarke, B., Eds.; Australian Centre for International Agricultural Research: Canberra, Australia, 2018; pp. 147–151.
7. Dahler, J.M.; McConchie, C.; Turnbull, C.G.N. Quantification of cyanogenic glycosides in seedlings of three *Macadamia* (Proteaceae) species. *Aust. J. Bot.* **1995**, *43*, 619–628. [[CrossRef](#)]
8. Vivian-Smith, A. Cyanogenesis and the Fruit Growth of Three *Macadamia* Species. Bachelor's Thesis, University of Queensland, Brisbane, Australia, 1995.
9. Gross, C.L.; Weston, P.H. *Macadamia janseni* (Proteaceae), a new species from central Queensland. *Aust. Syst. Bot.* **1992**, *5*, 725–728. [[CrossRef](#)]
10. Shigeura, G.T.; Ooka, H. *Macadamia* nuts in Hawaii, history and production. In *Research Series, Hawaii Institute of Tropical Agriculture and Human Resources*; University of Hawaii: Honolulu, HI, USA, 1984; pp. 6–22.
11. Hamilton, R.A.; Fukunaga, E.T. *Growing Macadamia Nuts in Hawaii*; Hawaii Agricultural Experiment Station: Honolulu, HI, USA, 1959; Volume 121, p. 59.
12. Topp, B.; Hardner, C.M.; Neal, J.; Kelly, A.; Russell, D.; McConchie, C.; O'hare, P. Overview of the Australian macadamia industry breeding program. *Acta Hortic.* **2016**, *1127*, 45–50. [[CrossRef](#)]
13. Yang, S.; Xue, S.; Kang, W.; Qian, Z.; Yi, Z. Genetic diversity and population structure of *Miscanthus lutarioriparius*, an endemic plant of China. *PLoS ONE* **2019**, *14*, e0211471. [[CrossRef](#)]
14. Singh, B.P.; Singh, B.; Mishra, S.; Kumar, V.; Singh, N.K. Genetic diversity and population structure in Indian wild rice accessions. *Aust. J. Crop Sci.* **2016**, *10*, 144–151.
15. Steiger, D.L.; Moore, P.H.; Zee, F.; Liu, Z.; Ming, R. Genetic relationships of macadamia cultivars and species revealed by AFLP markers. *Euphytica* **2003**, *132*, 269–277. [[CrossRef](#)]
16. Nock, C.J.; Hardner, C.M.; Montenegro, J.D.; Ahmad Termizi, A.A.; Hayashi, S.; Playford, J.; Edwards, D.; Batley, J. Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. *Front. Plant Sci.* **2019**, *10*, 1–15. [[CrossRef](#)]
17. Saeed, S.; Barozai, M.Y. A review on genetic diversity on wild plants by using different genetic markers. *Pure Appl. Biol.* **2012**, *1*, 68–71. [[CrossRef](#)]
18. Aradhya, M.; Yee, L.; Zee, F.; Manshardt, R. Genetic variability in *Macadamia*. *Genet. Resour. Crop Evol.* **1998**, *45*, 19–32. [[CrossRef](#)]
19. Mast, A.R.; Willis, C.L.; Jones, E.H.; Downs, K.M.; Weston, P.H. 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.* **2008**, *95*, 843–870. [[CrossRef](#)] [[PubMed](#)]
20. Neal, J.M. *The Impact of Habitat Fragmentation on Wild Macadamia Integrifolia Maiden and Betche (Proteaceae) Population Viability*; University of New England: Armidale, Australia, 2007.
21. O'Connor, K.; Powell, M.; Nock, C.; Shapcott, A. Crop to wild gene flow and genetic diversity in a vulnerable *Macadamia* (Proteaceae) species in New South Wales, Australia. *Biol. Conserv.* **2015**, *191*, 504–511. [[CrossRef](#)]
22. Spain, C.S.; Lowe, A.J. Genetic consequences of subtropical rainforest fragmentation on *Macadamia tetraphylla* (Proteaceae). *Silvae Genet.* **2011**, *60*, 241–249. [[CrossRef](#)]
23. Nock, C.J.; Elphinstone, M.S.; Ablett, G.; Kawamata, A.; Hancock, W.; Hardner, C.M.; King, G.J. Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild *Macadamia* (Proteaceae). *Appl. Plant Sci.* **2014**, *2*, apps.1300089. [[CrossRef](#)] [[PubMed](#)]
24. Roorikwal, M.; von Wettberg, E.J.; Upadhyaya, H.D.; Warschefsky, E.; Rathore, A.; Varshney, R.K. Exploring germplasm diversity to understand the domestication process in *Cicer* spp. using SNP and DArT markers. *PLoS ONE* **2014**, *9*, e102016. [[CrossRef](#)] [[PubMed](#)]
25. Baloch, F.S.; Alsaleh, A.; Shahid, M.Q.; Çiftçi, V.; de Miera, L.E.S.; Aasim, M.; Nadeem, M.A.; Aktaş, H.; Özkan, H.; Hatipoğlu, R. A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from Central Fertile Crescent. *PLoS ONE* **2017**, *12*, e0167821. [[CrossRef](#)] [[PubMed](#)]
26. Grzebelus, D.; Iorizzo, M.; Senalik, D.; Ellison, S.; Cavagnaro, P.; Macko-Podgorni, A.; Heller-Uszynska, K.; Kilian, A.; Nothnagel, T.; Allender, C.; et al. Diversity, genetic mapping, and signatures of domestication in the carrot (*Daucus carota* L.) genome, as revealed by Diversity Arrays Technology (DArT) markers. *Mol. Breed.* **2014**, *33*, 625–637. [[CrossRef](#)]
27. Alam, M.; Neal, J.; O'Connor, K.; Kilian, A.; Topp, B. Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS ONE* **2018**, *13*, e0203465. [[CrossRef](#)]

28. O'Connor, K.; Kilian, A.; Hayes, B.; Hardner, C.; Nock, C.; Baten, A.; Alam, M.; Topp, B. Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDArT markers. *Tree Genet. Genomes* **2019**, *15*, 24. [[CrossRef](#)]
29. Sharp, D.; Playford, J. Native macadamia germplasm resources. In *Australian Macadamia Society Annual Conference Report*; Australian Macadamia Society: Lismore, Australia, 1997; p. 20.
30. Peace, C.; Vithanage, V.; Turnbull, C.; Carroll, B.J. Characterising macadamia germplasm with codominant radiolabelled DNA fingerprinting (RAF) markers. *Acta Horticolt.* **2002**, *575*, 371–380. [[CrossRef](#)]
31. Storey, W.B. The ternifolia group of *Macadamia* species. *Pac. Sci.* **1965**, *19*, 507–514.
32. Peace, C.; Vithanage, V.; Neal, J.; Turnbull, C.; Carroll, B. A comparison of molecular markers for genetic analysis of macadamia. *J. Horticolt. Sci. Biotechnol.* **2004**, *79*, 965–970. [[CrossRef](#)]
33. Sun, M. Effects of population size, mating system, and evolutionary origin on genetic diversity in *Spiranthes sinensis* and *S. hongkongensis*. *Conserv. Biol.* **1996**, *10*, 785–795. [[CrossRef](#)]
34. Shapcott, A.; Lamont, R.W.; O'Connor, K.M.; James, H.; Conroy, G.C. Population genetics of *Philotheca sporadica* (Rutaceae) to advise an offset translocation program. *Conserv. Genet.* **2015**, *16*, 687. [[CrossRef](#)]
35. Sniderman, J.K.; Jordan, G.J.; Cowling, R.M. Fossil evidence for a hyperdiverse sclerophyll flora under a non-Mediterranean-type climate. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3423–3428. [[CrossRef](#)]
36. Brunet, J.; Larson-Rabin, Z.; Stewart, C.M. The distribution of genetic diversity within and among populations of the rocky mountain Columbine: The impact of gene flow, pollinators, and mating system. *Int. J. Plant Sci.* **2012**, *173*, 484–494. [[CrossRef](#)]
37. Pisanu, P.C. Survivorship of the Threatened Subtropical Rainforest Tree *Macadamia tetraphylla* L. Johnson (Proteaceae) in Small Habitat Fragments. Ph.D. Thesis, University of New England, Armidale, Australia, 2001.
38. Halász, J.; Kodad, O.; Galiba, G.; Skola, I.; Ercisli, S.; Ledbetter, C.; Hegedűs, A. Genetic variability is preserved among strongly differentiated and geographically diverse almond germplasm: An assessment by simple sequence repeat markers. *Tree Genet. Genomes* **2019**, *15*, 1–13. [[CrossRef](#)]
39. Nkatha, G.M.; Alice, M.; Anmin, Y.; Jianchu, X.; Aizhong, L. Genetic diversity analysis reveals genetic differentiation and strong population structure in *Calotropis* plants. *Sci. Rep.* **2018**, *8*, 1–10.
40. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C. Diversity arrays technology: A generic genome profiling technology on open platforms. In *Data Production and Analysis in Population Genomics: Methods and Protocols*; Humana Press: Totowa, NJ, USA, 2012; pp. 67–89.
41. Pritchard, J.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
42. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **2003**, *164*, 1567–1587.
43. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* **2007**, *7*, 574–578. [[CrossRef](#)]
44. Hubisz, M.J.; Falush, D.; Stephens, M.; Pritchard, J.K. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **2009**, *9*, 1322–1332. [[CrossRef](#)]
45. Earl, D.; von Holdt, B. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [[CrossRef](#)]
46. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)]
47. Bernard, A.; Barreneche, T.; Lheureux, F.; Dirlwanger, E. Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLoS ONE* **2018**, *13*, e0208021. [[CrossRef](#)]
48. Peakall, R.; Smouse, P. GenALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *28*, 2537–2539. [[CrossRef](#)]
49. Nei, M. Genetic distance between populations. *Am. Nat.* **1972**, *106*, 283–292. [[CrossRef](#)]
50. Perrier, X.; Jacquemoud-Collet, J.P. DARwin Software. 2006. Available online: <http://darwin.cirad.fr/darwin> (accessed on 3 June 2020).

