



OPEN ACCESS

EDITED BY
MANUEL TALÓN,
Instituto Valenciano de Investigaciones
Agrarias, Spain

REVIEWED BY
Priyanka Sharma,
University of Queensland, Australia
Alberto García,
Universitat Politècnica de València,
Spain

*CORRESPONDENCE
Jun Ni
nijun@gxu.edu.cn
Zeng-Fu Xu
zfxu@gxu.edu.cn

SPECIALTY SECTION
This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

RECEIVED 30 July 2022
ACCEPTED 11 October 2022
PUBLISHED 27 October 2022

CITATION
Wang P, Mo Y, Wang Y, Fei Y, Huang J,
Ni J and Xu Z-F (2022) Macadamia
germplasm and genomic database
(MacadamiaGGD): A comprehensive
platform for germplasm innovation
and functional genomics
in *Macadamia*.
Front. Plant Sci. 13:1007266.
doi: 10.3389/fpls.2022.1007266

COPYRIGHT
© 2022 Wang, Mo, Wang, Fei, Huang, Ni
and Xu. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Macadamia germplasm and genomic database (MacadamiaGGD): A comprehensive platform for germplasm innovation and functional genomics in *Macadamia*

Pan Wang^{1,2}, Yi Mo^{1,2}, Yi Wang^{1,2}, Yuchong Fei^{1,2},
Jianting Huang^{1,2}, Jun Ni^{1,2*} and Zeng-Fu Xu^{1,2*}

¹State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Forestry, Guangxi University, Nanning, China, ²Key Laboratory of National Forestry and Grassland Administration for Fast-Growing Tree Breeding and Cultivation in Central and Southern China, College of Forestry, Guangxi University, Nanning, China

As an important nut crop species, macadamia continues to gain increased amounts of attention worldwide. Nevertheless, with the vast increase in macadamia omic data, it is becoming difficult for researchers to effectively process and utilize the information. In this work, we developed the first integrated germplasm and genomic database for macadamia (MacadamiaGGD), which includes five genomes of four species; three chloroplast and mitochondrial genomes; genome annotations; transcriptomic data for three macadamia varieties, germplasm data for four species and 262 main varieties; nine genetic linkage maps; and 35 single-nucleotide polymorphisms (SNPs). The database serves as a valuable collection of simple sequence repeat (SSR) markers, including both markers that are based on macadamia genomic sequences and developed in this study and markers developed previously. MacadamiaGGD is also integrated with multiple bioinformatic tools, such as search, JBrowse, BLAST, primer designer, sequence fetch, enrichment analysis, multiple sequence alignment, genome alignment, and gene homology annotation, which allows users to conveniently analyze their data of interest. MacadamiaGGD is freely available online (<http://MacadamiaGGD.net>). We believe that the database and additional information of the SSR markers can help scientists better understand the genomic sequence information of macadamia and further facilitate molecular breeding efforts of this species.

KEYWORDS

Macadamia, germplasm, genome, SSR, SNP, MacadamiaGGD, genomic database, molecular breeding

Introduction

Macadamia (*Macadamia* spp.), which belongs to the Proteaceae family (Urata, 1954), is an evergreen perennial flowering plant species (Storey and Hamilton, 1953) originating from southern Queensland and northern New South Wales in Australia (Moncur et al., 1985). Macadamia has already become one of the most important economic oil crop species worldwide (Sedgley, 1983; Aradhya et al., 1998; Topp et al., 2019) due to the high level of monounsaturated fatty acid-palmitoleic acid (omega-7) in its nuts, which can effectively lower blood total cholesterol and benefit human health (Nagao et al., 1992; Moodley et al., 2007; Arroyo-Caro et al., 2016). To date, four macadamia species, namely, *Macadamia integrifolia* (Maiden & Betche), *M. tetraphylla* (L. A. S. Johnson), *M. ternifolia* (F. Muell), and *M. janseni* (C.L. Gross & P.H. Weston), have been identified (Mast et al., 2008), among which only *M. integrifolia*, *M. tetraphylla*, and their hybrids are most widely planted worldwide (SAMAC, 2020). The other two species, *M. ternifolia* and *M. janseni*, have not yet been used for any commercial purpose because they produce only small, unpalatable, bitter, inedible nuts, the mature nuts of which contain high cyanogenic glycoside levels (Trueman, 2013; Mai et al., 2020).

Macadamia plants are diploid ($2n = 28$) (Peace et al., 2003) and their genome size ranges from 758 to 896 megabase (Mb) (Nock et al., 2020; Niu et al., 2022a). In recent years, several *de novo*-assembled macadamia genomes have been reported, providing new insight for genetic breeding. In 2016, the first assembled draft genome of macadamia (*M. integrifolia* cultivar HAES 741) was finished and released by Nock's lab, the staff of whom used the short-read Illumina sequence platform (193493 scaffolds, N50 = 4745 bp, 518 Mb) (Nock et al., 2016). In 2020, the first sequence-based genetic linkage maps of macadamia were constructed (Langdon et al., 2020). In 2020, an improved chromosome-scale genome assembly of *M. integrifolia* cultivar HAES 741 was completed by the use of the short-read Illumina and long-read Pacific Biosciences (PacBio) sequencing platforms (4094 scaffolds, N50 = 413 kb, 745 Mb) (Nock et al., 2020). Furthermore, in 2020, by using the third-generation sequencing (TGS) platforms Oxford Nanopore (PromethION), PacBio (Sequel I), and BGI (Single-tube Long Fragment Read), researchers assembled the genome of *M. janseni* (Murigneux et al., 2020). In addition, the genomes of *M. integrifolia* (249 contigs, N50 = 5.3 Mb, 738 Mb), *M. tetraphylla* (153 contigs, N50 = 10.0 Mb, 707 Mb), *M. ternifolia* (211 contigs, N50 = 6.4 Mb, 716 Mb), and *M. janseni* (284 contigs, N50 = 4.5 Mb, 738 Mb) were assembled by use of the PacBio HiFi TGS platform (Sharma et al., 2021a). The genome of *M. janseni* has been improved by Hi-C assembly (219 scaffolds, N50 = 52 Mb, 758 Mb) (Sharma et al., 2021c) and was further updated by the latest hifiasm assembly (779 contigs, N50 = 46 Mb, 826 Mb) (Sharma et al., 2021b). Recently, the genome of the cultivar HAES 344 was sequenced and assembled into 14 pseudochromosomes by the use of Illumina NovaSeq and PacBio Sequel II sequencing (5387

contigs, N50 = 281 kb, 794 Mb) (Lin et al., 2022). A chromosome-scale genome assembly of *M. tetraphylla* has also been constructed from long-read Oxford Nanopore Technologies (ONT) sequencing data (1059 scaffolds, N50 = 51 Mb, 751 Mb) (Niu et al., 2022a). Moreover, in recent years, the chloroplast and mitochondrion genomes of *M. integrifolia*, *M. tetraphylla*, and *M. ternifolia* have been assembled and thoroughly annotated (Niu et al., 2022b).

As inbreeding decline occurs in macadamia, it is vitally important to understand the genetic distances between individuals (Steiger et al., 2003). The morphological characteristics of macadamia could be greatly influenced by the environment; thus, it is sometimes difficult to identify genetic relationships through phenotypic observations (Hardner, 2016). The use of DNA marker systems has become one of the most efficient strategies to evaluate genetic distance and genetic foundation (Ranketse et al., 2022). DNA marker systems, including isozyme (Vithanage and Winks, 1992; Aradhya et al., 1998), randomly amplified DNA fingerprinting (RAF) (Peace et al., 2002; Peace et al., 2004; Peace et al., 2005), amplified fragment length polymorphism (AFLP) (Steiger et al., 2003), sequence tagged site (STS) (Vithanage et al., 1998), random amplified polymorphic DNA (RAPD) (Vithanage et al., 1998), randomly amplified microsatellite fingerprinting (RAMiFi) (Peace et al., 2004), simple sequence repeat (SSR) (Schmidt et al., 2006; Nock et al., 2014b; Langdon et al., 2019; Ranketse et al., 2022), diversity array technology (DART) and single-nucleotide polymorphism (SNP) markers (Alam et al., 2018; O'Connor et al., 2019b), have been developed for the genetic and molecular breeding of macadamia. Genome-wide association studies (GWASs) have also greatly facilitated the identification of new molecular markers associated with yield traits (O'Connor et al., 2019a; O'Connor et al., 2020). As codominant, highly reproducible, highly polymorphic and cost-efficient DNA markers, SSRs have been preferred for use in studies of genetic identification and diversity analysis. To date, although the sequencing of the whole genomes of different macadamia species has been completed, genome-based development of SSR markers has not been reported.

With the rapidly developed sequencing technologies, the genomes of dozens of plant species have been sequenced each year. Nevertheless, how to integrate and well manage the large amount of omics data is still a task. In recent years, the genomic databases of some economic crops were well constructed and greatly facilitated the researchers to use the genome, transcriptome, or phenotype data. Citrus Genome Database (CGD, <https://www.citrusgenomedb.org/>) integrates genomes, maps, markers, phenotype data, and quantitative trait loci of agronomic traits of 25 citrus species. The Rice Genome Hub (RGH, <https://rice-genome-hub.southgreen.fr/>), which is part of the South Green Bioinformatics platform, also integrates large amount of rice omics data with a large number of powerful in-house tools (Droc et al., 2019). Rice Annotation Project Database

(RAP-DB, <https://rapdb.dna.affrc.go.jp/>) is consisted of updated genome annotation and focuses on the comprehensive analysis of genome structure and function of rice genes (Project, 2007). Gossypium Resource and Network Database (GRAND, <http://grand.cricaas.com.cn>) contains the genomic, transcriptomic, phenotypic, and integrative analysis tools for cotton (Zhang et al., 2022). With the inspirations from these databases, in this study we developed the first integrated germplasm and functional genomic database for macadamia (MacadamiaGGD).

Currently, large amounts of macadamia omics data lack centralized management. These data are distributed across multiple repositories or personal websites, with the same data from the same source in different repositories. In addition, many macadamia omics data lack the management of versions. The same data has different versions and accession numbers in different repositories, which can make it difficult for users to find the most updated dataset. The main purpose of the MacadamiaGGD described in this article is to provide the germplasm data, genome resources, transcriptome (RNA-seq) data, molecular marker information and genetic linkage map information to assist in the scientific research and molecular breeding of macadamia. And several commonly used bioinformatics tools are also integrated with MacadamiaGGD, which can help the researchers better utilize the database.

Materials and methods

Data sources and processing

In MacadamiaGGD, we integrated the genetic information data, including that of five genomes of four species, the chloroplast and mitochondrion genomes of three species and genome annotations, which were previously released in public databases, including the National Center for Biotechnology Information (NCBI) Assembly database, the *GigaScience* database (*GigaDB*), and the China National Center for Bioinformation (CNCB) Genome Warehouse (GWH) database. In addition, transcriptomic data for three macadamia varieties were downloaded from the NCBI Sequence Read Archive (SRA) database. The germplasm, genetic linkage map, SNP and SSR marker data were retrieved from the NCBI PubMed database and other databases, as summarized in Table 1. The components of data integration mainly include the data source, the data transform, and the data sink in the database. Extract, transform, and load (ETL) architecture was applied to data integration. In data integration process, raw data were collected, transformed, sorted, cleaned, aggregated, and stored *via* using PostgreSQL 9.5.25, Scala 2.13.1, AKKA 2.6.5, and SBT 1.3.5 (Figures 1A, B). Processed raw data were applied for variation calling and data visualization though using HTML5, CSS3, Java Script, Slick 3.3.2, Bootstrap 3.3.0 and Play Framework 2.8.2 (Figure 1B).

Development of the database

MacadamiaGGD was deployed in the Ubuntu 16.04 operation system using AKKA 2.6.5 (<https://akka.io>) as the web server, PostgreSQL 9.5.25 (<https://www.postgresql.org>) as the database server, Scala 2.13.1 (<https://www.scala-lang.org>) as the programming language and SBT 1.3.5 (<https://www.scala-sbt.org>) as the interactive building tool. All the data were managed and stored in the PostgreSQL Database. The website interface was generated *via* Bootstrap 3.3.0 (<https://getbootstrap.com>) and Play Framework 2.8.2 (<https://www.playframework.com/>). The web interface of MacadamiaGGD was developed using HTML5, CSS3, Java Script. The query function was enforced based on the Slick 3.3.2 middleware tier. JBrowse 1.16.6 (<https://www.jbrowse.org>) was used for genome visualization.

Sample collection and DNA isolation

Leaf samples of 21 macadamia accessions for DNA isolation were collected from the macadamia plantation in Chongzuo, Guangxi, China (Table S1). The DNA was isolated following a previously described method (Doyle, 1991), with slight modifications. To avoid problems of low efficiency and insufficient grinding due to manual grinding, young leaves were ground in a Tissuelyser-192 (Shanghai Jingxin Industrial Development Co., Ltd., China) and extracted with a 2% cetyltrimethylammonium bromide (CTAB) buffer. Nucleic acids were isolated with a chloroform: isoamyl alcohol (24:1) solution. DNA was purified with ethanol and resuspended in sterile distilled water. The DNA quality and concentration were assessed using ultraviolet spectrometry *via* a Nanodrop 2000c (Thermo Fisher Scientific, MA, USA) and agarose gel electrophoresis. The purified DNA was stored at -20°C until use.

Genome-wide SSR screening and characterization

New microsatellite markers were screened in the *M. integrifolia* HAES 741 reference genome (<https://www.ncbi.nlm.nih.gov/bioproject/748012>) by using SSRHunter 1.3 (<http://www.bio2soft.net>) (Li and Wan, 2005). The search criteria were set as 2, 3, and 4 nucleotides, corresponding to at least 4 repetitions. Afterward, the SSRs, comprising no fewer than 30 repeated motifs and being evenly distributed on each chromosome, were preferentially selected. To further confirm the quality of the SSR markers, each sequence was again queried *via* BLAST within MacadamiaGGD and tested *via* polymerase chain reaction (PCR).

Primer 3 (<https://primer3.org>) was used to design primer pairs flanking the sequences of the screened SSR motifs. The primer design parameters were as follows: primer length, 17-25

TABLE 1 Summary of all datasets in MacadamiaGGD.

Dataset	Species	References	Repository/ Accession number	URL
Germplasm	<i>M. integrifolia</i> <i>M. ternifolia</i> <i>M. tetraphylla</i> <i>M. janseni</i>	Vithanage and Winks (1992); Aradhya et al. (1998); Peace et al. (2002); Peace et al. (2005); Allan (2007); He (2008); Gitonga et al. (2009); Hardner et al. (2009); Machado Neto and Moryia (2010); Zhang (2011); Hardner (2016); Zeng and Du (2017); Alam et al. (2018); Tang et al. (2018); Toft et al. (2018); Langdon et al. (2019); O'Connor et al. (2019b); Tan et al. (2019); Tan et al. (2020); Mai et al. (2021); Tan et al. (2021); Lin et al. (2022)		http://MacadamiaGGD.net/nut/toRef
Genome assembly	<i>M. integrifolia</i> HAES 741	Nock et al. (2020)	NCBI/PRJNA748012	https://www.ncbi.nlm.nih.gov/bioproject/748012
	<i>M. integrifolia</i> HAES 344	Lin et al. (2022)	CNCB/PRJCA004595	https://ngdc.cncb.ac.cn/gwh/Assembly/23196/show
	<i>M. tetraphylla</i>	Sharma et al. (2021a)	GigaDB/100906; NCBI/PRJNA694456	http://gigadb.org/dataset/view/id/100906/
	<i>M. ternifolia</i>	Sharma et al. (2021a)	GigaDB/100906; NCBI/PRJNA694456	http://gigadb.org/dataset/view/id/100906/
	<i>M. janseni</i>	Sharma et al. (2021a)	GigaDB/100906; NCBI/PRJNA694456	http://gigadb.org/dataset/view/id/100906/
Genome annotation	<i>M. integrifolia</i> HAES 741	Nock et al. (2020)	NCBI/PRJNA748012	https://www.ncbi.nlm.nih.gov/bioproject/748012
	<i>M. integrifolia</i> HAES 344	Lin et al. (2022)	CNCB/PRJCA004595	https://ngdc.cncb.ac.cn/gwh/Assembly/23196/show
Chloroplast assembly and annotation	<i>M. integrifolia</i>	Nock et al. (2014a)	NCBI/PRJNA264682	https://www.ncbi.nlm.nih.gov/genome/?term=txid60698
	<i>M. ternifolia</i>	Liu et al. (2017)	NCBI/PRJNA421511	https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/66349/
	<i>M. tetraphylla</i>	Liu et al. (2018)	NCBI/MH778649	https://www.ncbi.nlm.nih.gov/nucleotide/MH778649
Mitochondrion assembly and annotation	<i>M. integrifolia</i>	Niu et al. (2022b)	NCBI/MW566570	https://www.ncbi.nlm.nih.gov/nucleotide/MW566570
	<i>M. ternifolia</i>	Niu et al. (2022b)	NCBI/MW566571	https://www.ncbi.nlm.nih.gov/nucleotide/MW566571
	<i>M. tetraphylla</i>	Niu et al. (2022b)	NCBI/MW566572	https://www.ncbi.nlm.nih.gov/nucleotide/MW566572
Transcriptome	<i>M. integrifolia</i> HAES 741	Nock et al. (2020)	NCBI/PRJNA593881	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA593881
	<i>M. integrifolia</i> HAES 344	Lin et al. (2022)	NCBI/PRJNA706119	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA706119
	<i>M. integrifolia</i> H2	Lin et al. (2022)	NCBI/PRJNA706119	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA706119
Genetic linkage maps	<i>M. integrifolia</i>	Langdon et al. (2020)		https://researchportal.scu.edu.au/esploro/outputs/dataset/991012821025202368
SSRs	<i>M. integrifolia</i>	Schmidt et al. (2006); Nock et al. (2014b); Langdon et al. (2019)		http://MacadamiaGGD.net/nut/toRef
SNPs	<i>M. integrifolia</i> <i>M. integrifolia</i> <i>M. ternifolia</i> <i>M. integrifolia</i> × <i>M. tetraphylla</i> <i>M. janseni</i>	O'Connor et al. (2019a); O'Connor et al. (2020) (Alam et al., 2018)		http://MacadamiaGGD.net/nut/toRef

The datasets were deposited in the repositories of the China National Center for Bioinformatics (CNCB), the National Center for Biotechnology Information (NCBI), and the GigaDB.

bp; melting temperature (T_m), 53 °C; amplicon size, 350-500 bp; and GC content, 40-60%.

Marker analysis, data analysis and map construction

The SSR PCR mixture (10 µL) comprised 1 µL of DNA, 0.4 µL of each primer (10 µM), 5 µL of Rapid Taq Master Mix (Vazyme, China) and 3.2 µL of double-distilled water. The amplification reaction program was as follows: 5 min at 95°C; 36 cycles of (30 s at 95°C, 53°C and 72°C); and a final extension of 5 min at 72 °C. Afterward, the mixture was held at 16°C. The PCR products were examined by electrophoresis on a 7% nondenaturing polyacrylamide gel run at 220 V for 40 min and visualized by silver staining. The density distribution map of polymorphic SSR markers on chromosomes was generated using MG2C software (http://mg2c.iask.in/mg2c_v2.1/).

Results

Overview of MacadamiaGGD

MacadamiaGGD contains the most comprehensive bioinformatics datasets of macadamia (including five genomes, a total of 89.28 Gb of transcriptomic data, three chloroplast and mitochondrion genomes, germplasm data for four species and 262 main varieties, nine genetic linkage maps, 35 SNPs and 657 SSR markers), which provides convenient access to the large amount of germplasm and genomic information of macadamia (Figure 1A). MacadamiaGGD is composed of 11 main functional modules: Home, Germplasm, Genomes, Expression, BLAST, Markers, Maps, Tools, References, Download and Help (Figure 1C). MacadamiaGGD can be used to search and visualize genomic information by using various tools, including search, JBrowse, BLAST, primer designer, sequence fetch, enrichment analysis, multiple sequence alignment, genome alignment, and gene homology annotation (Figure 1). MacadamiaGGD also provides information about macadamia germplasm and genome-related references. In summary, researchers can use the above functional modules of the database to quickly acquire the germplasm and genomic information of macadamia.

Germplasm

In the Germplasm module of MacadamiaGGD, 23 agronomic traits of four species and 16 agronomic traits of 262 main varieties were carefully described, including tree vigor, leaf type, fruit shape, flower color, the early-bloom stage and full-boom stage, and others. Users can easily obtain information on

the morphological characteristics of four macadamia species and 262 varieties in the germplasm module. In addition, a phylogenetic analysis tool based on the results of Alam et al. (2018), which shows genetic distances between individuals genotypes, is provided in this module.

Genome browse and search

The MacadamiaGGD database provides public information on the assembled genomes of the *M. integrifolia*, *M. tetraphylla*, *M. ternifolia*, and *M. janseni*, which are available in different public databases. For example, when “Genomes” is clicked on, the column header label appears, showing the suboptions as in Figure 2A. We can choose any label to access the sublinks and search for the needed information. When the user enters a gene “LOC122078696” in *Macadamia integrifolia* HAES 741 genome Browse, it will get the structure and function annotation information of all transcript of the gene (Figure 2B). Moreover, when the user clicks “Search”, a new layer appears with four options: “Keyword”, “Gene ID”, “Gene Name”, and “Region” (Figure 2C). Then, if one clicks “Gene ID”, the interface appears as a blank box (Figure 2C). The user can enter the gene “LOC122078696” in the box and click the Search button; then, the requested information is displayed (Figure 2C).

Genome JBrowse

Gene annotations in MacadamiaGGD are displayed graphically in the genome JBrowse, which includes the information of the gene location, nucleotide sequences, amino acid sequences, and other features. For example, if a user selects the genomic region from 192751 bp to 203445 bp on Chromosome 14 (NC_056557.1) for browsing, all genes located within this zone are displayed properly (Figure 2D). Further, when the mRNA XM_042644823.1 is clicked on, detailed information on its mRNA, coding sequence (CDS), and other features are displayed (Figure 2E).

Transcriptomes of macadamia from different tissues

In the expression module of MacadamiaGGD, a total of 89.28 Gb of raw RNA-seq data were collected from tissues of young leaves, shoots, and flowers from the cultivar ‘Mauka’ (Nock et al., 2020); tissues of leaves, stems, flowers, and roots from the cultivar ‘Kau’; and shells and kernels at five different development stages from cultivar ‘Hinde’ (Lin et al., 2022). By mapping the transcriptome data to the reference genome and using transcripts per million (TPM) for calculation, we acquired the expression matrix of the annotated genes of macadamia.

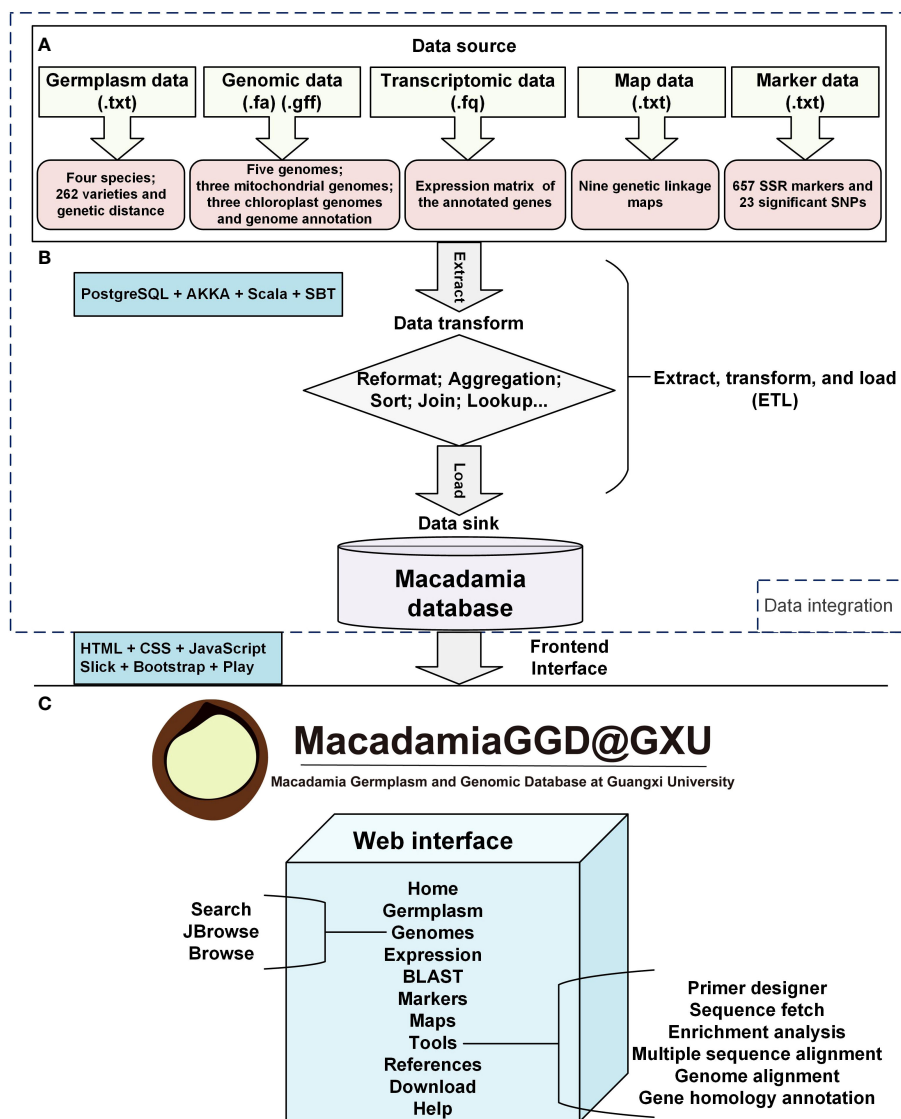


FIGURE 1

Feature diagram of MacadamiaGGD. MacadamiaGGD is a collection of germplasm, genomic, transcriptomic, maps, and molecular marker data of macadamia, and multiple bioinformatic tools. All the data are stored and managed in a PostgreSQL database. (A), Data source layer. (B), Middleware layer. (C), Application layer.

BLAST

BLAST is the most commonly used tool and is included as a separate module in the MacadamiaGGD database. It allows users to perform both BLASTp and BLASTn searches to rapidly align sequences to the database. In the BLAST module, pasting the DNA/protein sequences in the query box or uploading a FASTA file is acceptable. For example, the users can enter “Example 1” sequence in the blank box and select the against database type, e-value, and max target sequence number and then click the “Run” button to obtain the comparison results *via* the “BLASTp” function (Figure 3A). In addition, when pulling down the

search result interface, a user is presented with all the comparison results (Figure 3B), including the description information of the candidate subject sequences alignment parameters (Figure 3C) and the matching information between the query sequence and each subject sequence (Figure 3D).

Markers

In the “markers” module, we included 657 SSR markers and 35 SNPs. Macadamia trees have a relatively long juvenile period (commonly four to five years); thus, it would take a great deal of

time to select high-yielding cultivars for breeding. Molecular markers that are associated with key yield traits are extremely important for developing rapid cycle breeding programs in macadamia (O'Connor et al., 2020). To verify the polymorphism of SSR markers from previous research (Schmidt et al., 2006; Nock et al., 2014b; Langdon et al., 2019), we randomly selected 8 primer pairs from MacadamiaGGD (Table S2) and identified polymorphisms of these SSRs *via* electrophoresis. The results showed that the selected primer pairs were polymorphic.

In this study, a total of 145593 SSR loci were obtained from *M. integrifolia* HAES 741 genomic sequences (Nock et al., 2020). They were evenly distributed on 14 chromosomes, with an average density of 10400 loci per chromosome (Table 2). SSR motifs exist as one of three main types: dinucleotide repeats (DNRs), trinucleotide repeats (TNRs) and tetranucleotide repeats (TTRs). Among these SSRs, DNRs were the most abundant (115139), followed by TNRs (26400) and TTRs (4054), which accounted for 79%, 18% and 3%, respectively (Table 2). A total of 927 primer pairs were designed by the selection of the SSR loci with repeat numbers ≥ 30 from the total SSR loci (Table S3). Out of 927 amplified products, 605 primer pairs were polymorphic, with an average of 1.17 SSR markers per Mb on 14 chromosomes. According to the SSR density distribution map, chromosome 5 had the highest number of SSRs (81), but chromosome 12 had only 13 SSRs (Figure 4). In addition, a total of 35 SNPs were included in the “markers” module, which were significantly associated with the yield

component traits identified by genome-wide association studies (GWASs) (O'Connor et al., 2019a; O'Connor et al., 2020).

Maps

The map module contains nine genetic linkage maps derived from three macadamia cultivars, HAES 741, HVP A268 and HVP A4. In each map, there were 14 linkage groups (LGs), which correspond to the number of haploid chromosomes in macadamia. When the users open this module, the features of the maps are displayed, including the description and number of maps. The images of the maps appear at the lower left of the module, while the detailed information of the LG location, the marker numbers, the largest and smallest gap, the total length and the average length between markers is displayed at the lower right.

Tools

The tools module contains several utilities, including “Primer designer”, “Sequence Fetch”, “Enrichment analysis”, “Multiple sequence alignment”, “Genome alignment”, and “Gene homology annotation”, which allow a relatively complete bioinformatics analysis. The user can click the “Primer designer” button, input the nucleic acid sequence or

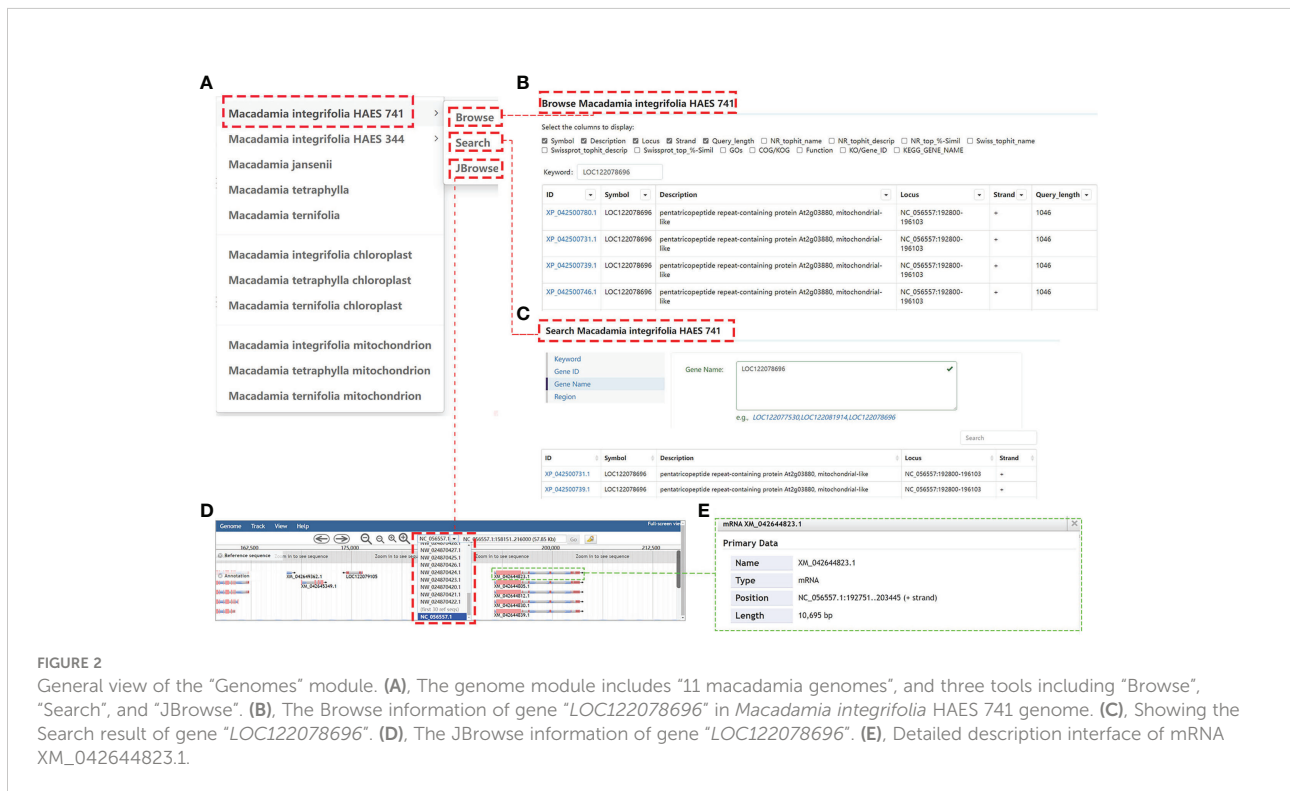


FIGURE 2 General view of the “Genomes” module. (A), The genome module includes “11 macadamia genomes”, and three tools including “Browse”, “Search”, and “JBrowse”. (B), The Browse information of gene “LOC122078696” in *Macadamia integrifolia* HAES 741 genome. (C), Showing the Search result of gene “LOC122078696”. (D), The JBrowse information of gene “LOC122078696”. (E), Detailed description interface of mRNA XM_042644823.1.

select a scaffold range, adjust the appropriate parameters, and click the “Run” button to obtain a satisfactory pair of primers. Users can screen functional genes of interest (GOIs) based on the data of the *M. integrifolia* transcriptome, click the “Enrichment analysis” button, input the gene ID in the dialog box and select Kyoto Encyclopedia of Genes and Genomes (KEGG) or Gene Ontology (GO) for functional clustering analysis. “Sequence Fetch” can be used to efficiently obtain the sequence of GOI from the *M. integrifolia* genome, which can acquire either a certain or multiple gene sequences at the same time. “Muscle” is a multisequence alignment tool that not only can be used to obtain homology between genes but also can be used to build an intuitive diagram. The “primer designer” tool can be used to design specific primers to clone GOIs for functional research. In addition, by using the “LASTZ” and “GeneWise” tools, users can complete genome alignment and gene homology annotation, respectively.

References

Currently, the “Reference” module contains the macadamia germplasm and genome-related references, which allows users to query approximately 40 articles information related to the data contained in MacadamiaGGD. The completion and optimization of macadamia genome sequencing results among these publications contribute to

the study of macadamia functional genomics and comparative genomics and are convenient for molecular plant breeding efforts.

A case study involving the use of MacadamiaGGD

MacadamiaGGD integrates BLAST, enrichment analysis, and other tools for functional genomic research of Macadamia. Acyltransferases are the potential molecular targets for genetic engineering to increase the oil content and alter the fatty acid composition in the oil crops (Zhang et al., 2021). Here, we provide a case study on the diacylglycerol acyltransferases (DGATs) of *M. integrifolia* by using the “BLAST”, “GO enrichment”, “JBrowse”, and “Gene Expression” function of MacadamiaGGD. By using BLAST in MacadamiaGGD, the Conserved Domains Database (CDD) of the NCBI database, the SMART database (<https://smart.embl.de/>) and MEGA 11 software (<https://megasoftware.net/>), we obtain one DGAT1 (MiDGAT1), three DGAT2 (MiDGAT2-1, MiDGAT2-2, MiDGAT2-3), and one DGAT3 (MiDGAT3) (Figure 5A). Of the five *MiDGAT* genes, two genes (*MiDGAT2-1*, *MiDGAT2-3*) were mapped to chromosome 14, and their physical positions were very close (Figure 5B).

To verify the expression features of *MiDGATs* during triacylglycerol (TAG) biosynthesis, we downloaded the transcriptome expression data of *M. integrifolia* kernel

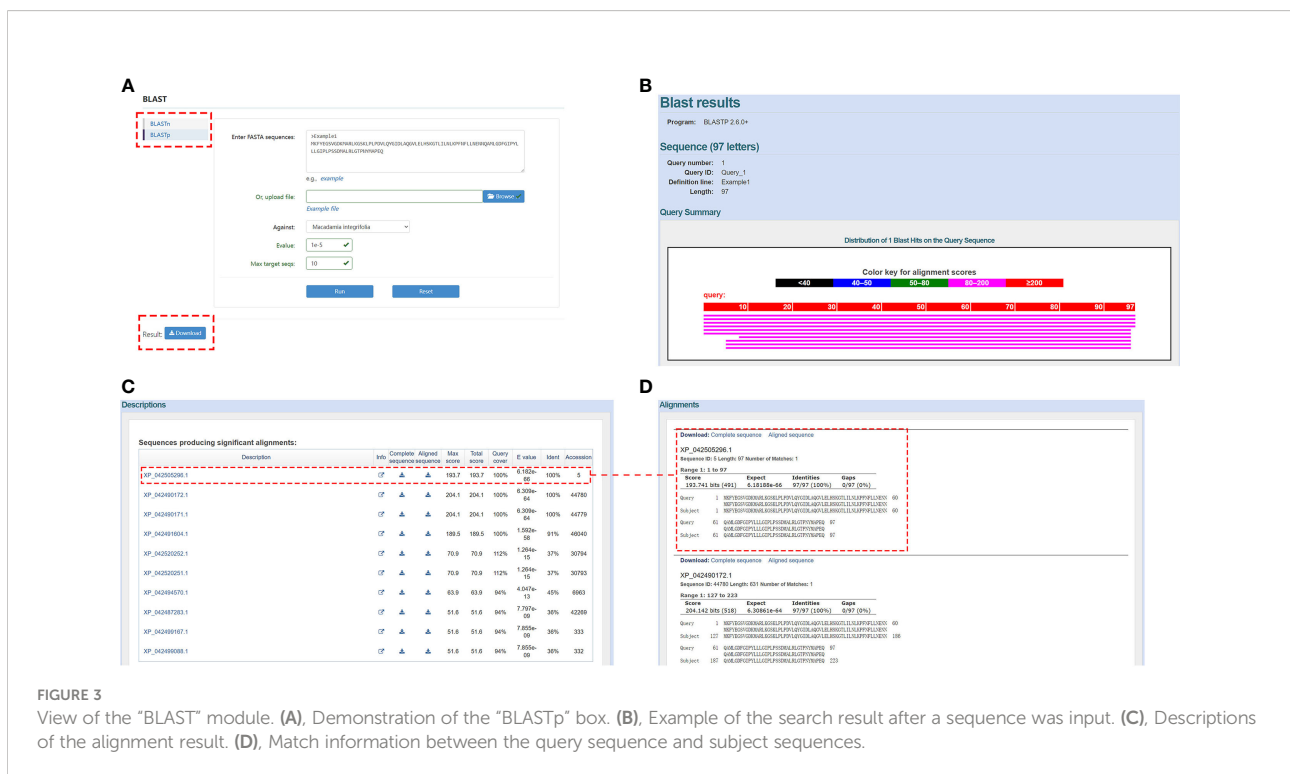


FIGURE 3 View of the “BLAST” module. (A), Demonstration of the “BLAST” box. (B), Example of the search result after a sequence was input. (C), Descriptions of the alignment result. (D), Match information between the query sequence and subject sequences.

TABLE 2 Characterization of the screened SSRs in *Macadamia integrifolia*.

Chromosome	DNR	TNR	TTR	All SSR loci	Proportion to all SSR loci (%)	All SSRs	SSRs densitydistribution on chromosome(1/Mb)
Chr1	6453	1276	196	7925	5.44	21	0.58
Chr2	10092	2386	352	12830	8.81	28	0.64
Chr3	9217	2048	349	11614	7.98	25	0.66
Chr4	8935	2142	318	11395	7.83	67	1.79
Chr5	11249	2537	400	14187	9.74	81	1.72
Chr6	8208	1823	291	10322	7.1	63	1.54
Chr7	8730	2065	289	11084	7.61	24	0.65
Chr8	8878	1991	288	11157	7.66	73	2.09
Chr9	6890	1566	249	8705	5.98	22	0.52
Chr10	7443	1833	269	9547	6.56	44	1.29
Chr11	7625	1856	271	9752	6.7	59	1.74
Chr12	7315	1666	250	9231	6.34	13	0.41
Chr13	6562	1516	257	8335	5.72	57	1.95
Chr14	7542	1695	275	9512	6.53	29	0.88
Total	115139	26400	4054	145593	100	606	

DNR, dinucleotide repeat; TNR, trinucleotide repeat; TTR, tetranucleotide repeat; Mb, megabase.

development from *MacadamiaGGD*. By using gene ontology (GO) annotation information available from *MacadamiaGGD*, we conducted the GO enrichment analysis of the five *MiDGAT* genes from *M. integrifolia*. The results showed the five *MiDGAT*s were enriched in more than 30 GO terms, which are involved in fatty acid and TAG biosynthesis in plants (Figure 5C). Further, we also investigated the expression profile of *MiDGAT*s at five stages of

kernel development. *MiDGAT2-1* and *MiDGAT2-3* were highly expressed in stages I and II (Figure 5D). *MiDGAT2-2* exhibited low expression levels in stages I and II, whereas it was highly expressed in stages III, IV, and V. Consistent with these results, The expression pattern of *MiDGAT2* was recently found to be mainly correlated with fatty acid biosynthesis at different stages of developing kernels (Gao et al., 2021).

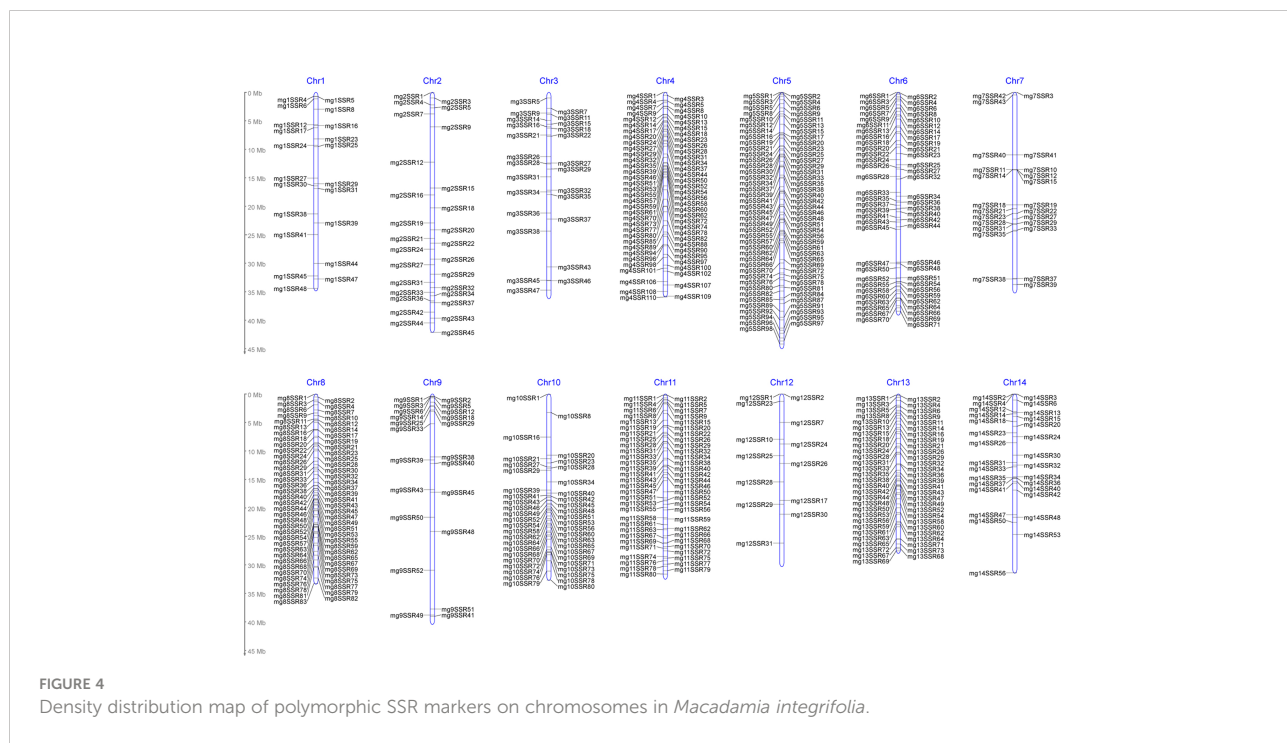
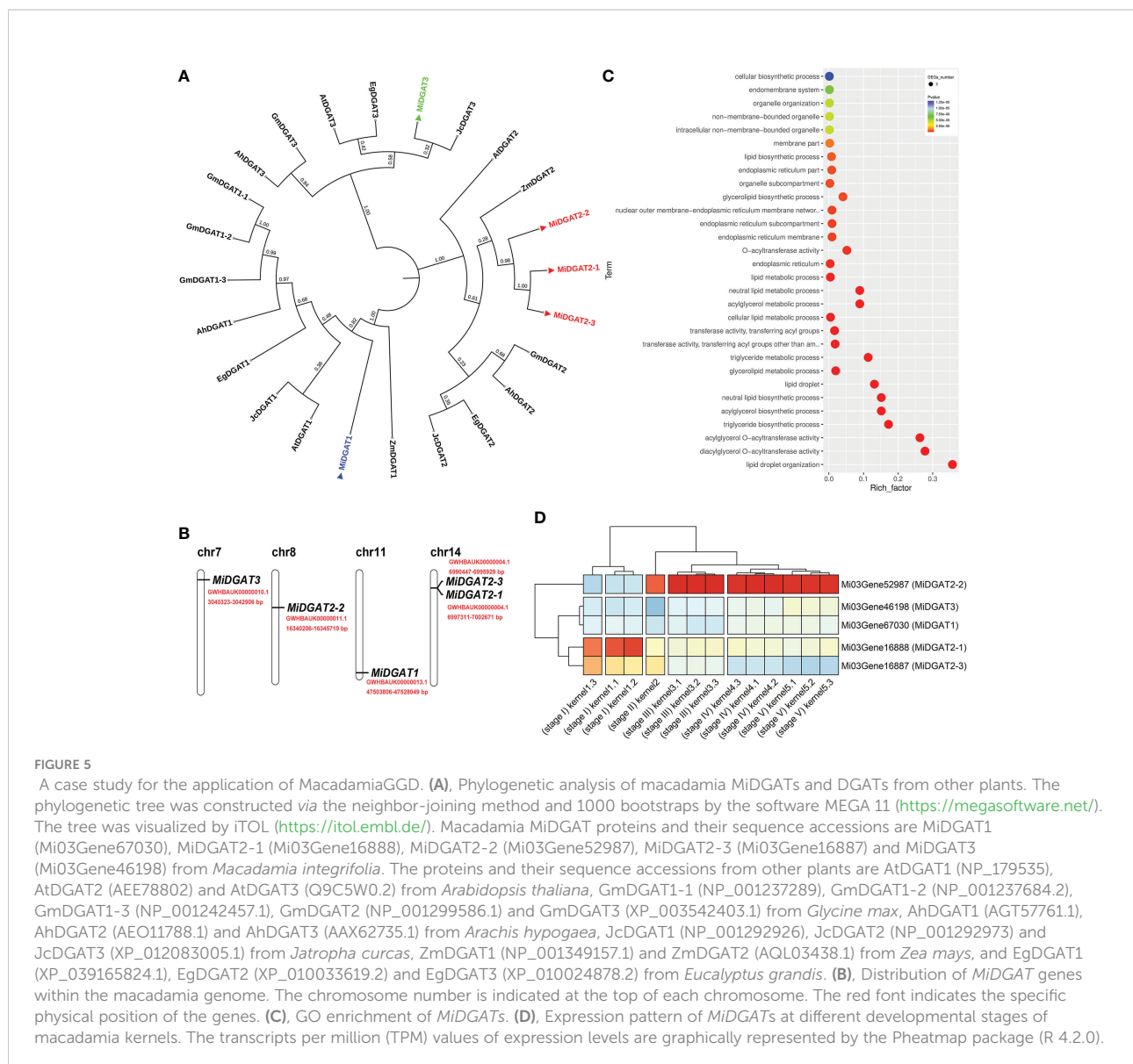


FIGURE 4 Density distribution map of polymorphic SSR markers on chromosomes in *Macadamia integrifolia*.



Discussion

The macadamia database MacadamiaGGD serves as an integrated germplasm and genomic research platform that can facilitate the genomic research and molecular breeding of macadamia. MacadamiaGGD integrates the currently published macadamia datasets of genomes, genetic maps, molecular markers, and morphological data of four macadamia species. MacadamiaGGD consists of 11 functional modules: Home, Germplasm, Genomes, Expression, BLAST, Markers, Maps, Tools, References, Download and Help.

Compared to other existing genome databases, the MacadamiaGGD provides a more comprehensive database and tools to characterize germplasm and genes of macadamia species. For example, “Phylogenetic Analysis”, which is

integrated in the Germplasm module of MacadamiaGGD, was not included in the Citrus Genome Database (CGD, <https://www.citrusgenomedb.org/>), the Rice Genome Hub (RGH, <https://rice-genome-hub.southgreen.fr>) (Droc et al., 2019), the Kiwifruit Genome Database (KGD; <http://kiwifruitgenome.org/>) (Yue et al., 2020), and the functional genomics database for cannabis (CannabisGDB, <https://gdb.supercann.net>) (Cai et al., 2021). Databases of two kinds of molecular markers, SSR and SNP, are included in MacadamiaGGD, but not available in RGH, KGD, CannabisGDB, and the Gossypium Resource and Network Database (GRAND, <http://grand.cricaas.com.cn>) (Zhang et al., 2022). And MacadamiaGGD provides genetic linkage maps of nine genotypes, whereas genetic linkage maps are not available in KGD, CannabisGDB, and GRAND. Given the comprehensive information, interactive nature, and user-

friendly database, MacadamiaGGD makes it easy to retrieve genomic information of macadamia. Thus, MacadamiaGGD not only provides a convenient way for researchers to understand and acquire basic germplasm and genomic information but also can largely help advance the molecular breeding of macadamia in the future.

The macadamia genome was used for the exploration of the SSR motifs, which were found to be evenly distributed across all 14 chromosomes. However, the percentage of the three SSR motifs was different, among which DNRs accounted for 79%, TNRs accounted for 18%, and TTRs accounted for 3%. This pattern is consistent with that in *Myrica rubra* (Jiao et al., 2012), in which DNRs were dominant. In this study, 927 primer pairs were designed for the verification of SSR locus polymorphisms, among which 605 primer pairs were found to be polymorphic. The density of microsatellite distribution was approximately 1.17 SSRs/Mb on 14 chromosomes, which was much higher than that in previous studies (Nock et al., 2014b). The main reason for this discrepancy may be due to the differences in genome quality and the SSR prediction method. In summary, we developed the first database of macadamia germplasm, genome, and genome-based SSR marker information, which will facilitate the molecular breeding of macadamia.

Conclusion

In conclusion, we developed the first comprehensive macadamia germplasm and genomic database MacadamiaGGD, which could serve as a central portal for macadamia species. MacadamiaGGD integrates data from germplasm, genomes, transcriptomes, genetic linkage maps, and SSR markers from various macadamia species. MacadamiaGGD also provides a group of user-friendly modules that enable users worldwide to efficiently retrieve and analyze genomic data. At present, MacadamiaGGD is in its first version but will be updated in a timely manner when new macadamia germplasm and omics data are available or published. We believe that MacadamiaGGD not only will broaden the understanding of the germplasm, genetics and genomics of macadamia species but also will facilitate the molecular breeding of macadamia.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

Z-FX and JN designed the research. PW, YM, YW, YF, and JH collected and processed genomic and germplasm data. YM and PW developed the SSRs. PW, Z-FX and JN wrote the first draft of this manuscript. All authors contributed to the edit of this manuscript and the construction of MacadamiaGGD. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by start-up research funds from the Guangxi University.

Acknowledgments

We would like to thank all the macadamia researchers who have created valuable data resources collected in MacadamiaGGD. Thanks to Mr. Zheng Yin and Mr. Zequn Zheng (VGsoft Team, China) for assisting in the website construction.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1007266/full#supplementary-material>

References

- Alam, M., Neal, J., O'Connor, K., Kilian, A., and Topp, B. (2018). Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS One* 13 (8), e0203465. doi: 10.1371/journal.pone.0203465
- Allan, P. (2007). Evaluation and identification of old and new macadamia cultivars and selections at pietermaritzburg. *South Afr. J. Plant Soil* 24 (2), 124–129. doi: 10.1080/02571862.2007.10634793
- Aradhya, M. K., Yee, L. K., Zee, F. T., and Manshardt, R. M. (1998). Genetic variability in macadamia. *Genet. Resour. Crop Evol.* 45 (1), 19–32. doi: 10.1023/A:1008634103954
- Arroyo-Caro, J. M., Manas-Fernandez, A., Alonso, D. L., and Garcia-Maroto, F. (2016). Type I diacylglycerol acyltransferase (MtDGAT1) from macadamia tetraphylla: cloning, characterization, and impact of its heterologous expression on triacylglycerol composition in yeast. *J. Agric. Food Chem.* 64 (1), 277–285. doi: 10.1021/acs.jafc.5b04805
- Cai, S., Zhang, Z., Huang, S., Bai, X., Huang, Z., Zhang, Y. J., et al. (2021). CannabisGDB: a comprehensive genomic database for *Cannabis sativa* L. *Plant Biotechnol. J.* 19 (5), 857–859. doi: 10.1111/pbi.13548
- Doyle, J. (1991). "DNA Protocols for plants," in *Molecular techniques in taxonomy*. Eds. G. M. Hewitt, A. W. B. Johnston and J. P. W. Young (Berlin, Heidelberg: Springer Berlin Heidelberg) 57, 283–293. doi: 10.1007/978-3-642-83962-7_18
- Droc, G., Dereeper, A., Ruiz, M., Antoine, C., Barca, M., and Tranchant-Dubreuil, C. (2019). The South green rice genome hub. *Plant and Animal Genome* 17, 1. doi: agritrop.cirad.fr/590957/
- Gao, Y., Sun, Y., Gao, H., Chen, Y., Wang, X., Xue, J., et al. (2021). Ectopic overexpression of a type-II DGAT (CeDGAT2-2) derived from oil-rich tuber of *Cyperus esculentus* enhances accumulation of oil and oleic acid in tobacco leaves. *Bioproducts* 14 (1), 76. doi: 10.1186/s13068-021-01928-8
- Gitonga, L. N., Muigai, A. W. T., Kahangi, E. M., and Ngamau, K. (2009). Status of macadamia production in Kenya and the potential of biotechnology in enhancing its genetic improvement. *J. Plant Breed. Crop Sci.* 1 (3), 049–059. doi: 10.5897/jpbcs.9000122
- Hardner, C. M. (2016). Macadamia domestication in hawaii. *Genet. Resour. Crop Evol.* 63 (8), 1411–1430. doi: 10.1007/s10722-015-0328-1
- Hardner, C. M., Peace, C., Lowe, A. J., Neal, J., Pisanu, P., Powell, M., et al. (2009). "Genetic resources and domestication of macadamia," in J. Janick (Ed.) *Horticultural Reviews*, vol. 35, (Hoboken, New Jersey: John Wiley & Sons, Inc) pp. 1–125. doi: 10.1002/9780470593776.ch1
- He, X. (2008). *Studies on genetic diversity of macadamia (Macadamia spp.) germplasm resources* (Southwest University: Master).
- Jiao, Y., Jia, H., Li, X., Chai, M., Jia, H., Chen, Z., et al. (2012). Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13 (1), 201. doi: 10.1186/1471-2164-13-201
- Langdon, K. S., King, G. J., Baten, A., Mauleon, R., Bundock, P. C., Topp, B. L., et al. (2020). Maximising recombination across macadamia populations to generate linkage maps for genome anchoring. *Sci. Rep.* 10 (1), 5048. doi: 10.1038/s41598-020-61708-6
- Langdon, K. S., King, G. J., and Nock, C. J. (2019). DNA Paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. *Tree Genet. Genomes* 15 (2), 29. doi: 10.1007/s11295-019-1336-7
- Lin, J., Zhang, W., Zhang, X., Ma, X., Zhang, S., Chen, S., et al. (2022). Signatures of selection in recently domesticated macadamia. *Nat. Commun.* 13 (1), 242. doi: 10.1038/s41467-021-27937-7
- Liu, J., Niu, Y. F., Ni, S. B., He, X. Y., and Shi, C. (2017). Complete chloroplast genome of a subtropical fruit tree *Macadamia ternifolia* (Proteaceae). *Mitochondrial DNA Part B Resour.* 2 (2), 738–739. doi: 10.1080/23802359.2017.1390401
- Liu, J., Niu, Y. F., Ni, S. B., He, X. Y., Zheng, C., Liu, Z. Y., et al. (2018). The whole chloroplast genome sequence of *Macadamia tetraphylla* (Proteaceae). *Mitochondrial DNA B Part Resour.* 3 (2), 1276–1277. doi: 10.1080/23802359.2018.1532836
- Li, Q., and Wan, J. M. (2005). SSRHunter: development of a local searching software for SSR sites. *Yi Chuan = Hereditas* 27 (5), 808–810. doi: 10.16288/j.ycz.2005.05.024
- Machado Neto, N. B., and Moryia, A. T. (2010). Variability in *Macadamia integrifolia* by RAPD markers. *Crop Breed. Appl. Biotechnol.* 10 (3), 266–270. doi: 10.1590/s1984-70332010000300013
- Mai, T. T. P., Alam, M. M., Hardner, C. M., Henry, R. J., and Topp, B. L. (2020). Genetic structure of wild germplasm of macadamia: species assignment, diversity and phylogeographic relationships. *Plants (Basel)* 9 (6), 741. doi: 10.3390/plants9060714
- Mai, T. T. P., Hardner, C. M., Alam, M. M., Henry, R. J., and Topp, B. L. (2021). Phenotypic characterisation for growth and nut characteristics revealed the extent of genetic diversity in wild macadamia germplasm. *Agriculture* 11 (7), 680. doi: 10.3390/agriculture11070680
- 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 (7), 843–870. doi: 10.3732/ajb.0700006
- Moncur, M. W., Stephenson, R. A., and Trochoulis, T. (1985). Floral development of *Macadamia integrifolia* maiden & betche under Australian conditions. *Scientia Hort.* 27 (1), 87–96. doi: 10.1016/0304-4238(85)90058-5
- Moodley, R., Kindness, A., and Jonnalagadda, S. B. (2007). Elemental composition and chemical characteristics of five edible nuts (almond, Brazil, pecan, macadamia and walnut) consumed in Southern Africa. *J. Environ. Sci. Health Part B* 42, 585–591. doi: 10.1080/03601230701391591
- Murigneux, V., Rai, S. K., Furtado, A., Bruxner, T. J. C., Tian, W., Harliwong, I., et al. (2020). Comparison of long-read methods for sequencing and assembly of a plant genome. *Gigascience* 9 (12), 1–11. doi: 10.1093/gigascience/giaa146
- Nagao, M. A., Hirae, H. H., and Stephenson, R. A. (1992). Macadamia: cultivation and physiology. *Crit. Rev. Plant Sci.* 10 (5), 441–470. doi: 10.1080/07352689209382321
- Niu, Y., Li, G., Ni, S., He, X., Zheng, C., Liu, Z., et al. (2022a). The chromosome-scale reference genome of *Macadamia tetraphylla* provides insights into fatty acid biosynthesis. *Front. Genet.* 13, 835363. doi: 10.3389/fgene.2022.835363
- Niu, Y., Lu, Y., Song, W., He, X., Liu, Z., Zheng, C., et al. (2022b). Assembly and comparative analysis of the complete mitochondrial genome of three macadamia species (*M. integrifolia*, *M. ternifolia* and *M. tetraphylla*). *PLoS One* 17 (5), e0263545. doi: 10.1371/journal.pone.0263545
- Nock, C. J., Baten, A., Barkla, B. J., Furtado, A., Henry, R. J., and King, G. J. (2016). Genome and transcriptome sequencing characterises the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* 17 (1), 937. doi: 10.1186/s12864-016-3272-3
- Nock, C. J., Baten, A., and King, G. J. (2014a). Complete chloroplast genome of *Macadamia integrifolia* confirms the position of the gondwanan early-diverging eudicot family proteaceae. *BMC Genomics* 15 (9), S13. doi: 10.1186/1471-2164-15-S9-S13
- Nock, C. J., Baten, A., Mauleon, R., Langdon, K. S., Topp, B., Hardner, C., et al. (2020). Chromosome-scale assembly and annotation of the macadamia genome (*Macadamia integrifolia* HAES 741). *G3-Genes Genomes Genet.* 10 (10), 3497–3504. doi: 10.1534/g3.120.401326
- Nock, C. J., Elphinstone, M. S., Ablett, G., Kawamata, A., Hancock, W., Hardner, C. M., et al. (2014b). Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild macadamia (Proteaceae). *Appl. Plant Sci.* 2 (4), 1300089. doi: 10.3732/apps.1300089
- O'Connor, K., Hayes, B., Hardner, C., Alam, M., and Topp, B. (2019a). Selecting for nut characteristics in macadamia using a genome-wide association study. *Hortic. Sci.* 54 (4), 629–632. doi: 10.21273/hortsci13297-18
- O'Connor, K., Hayes, B., Hardner, C., Nock, C., Baten, A., Alam, M., et al. (2020). Genome-wide association studies for yield component traits in a macadamia breeding population. *BMC Genomics* 21 (1), 199. doi: 10.1186/s12864-020-6575-3
- O'Connor, K., Kilian, A., Hayes, B., Hardner, C., Nock, C., Baten, A., et al. (2019b). Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDArT markers. *Tree Genet. Genomes* 15 (2), 24. doi: 10.1007/s11295-019-1331-z
- Peace, C. P., Allan, P., Vithanage, V., Turnbull, C. N., and Carroll, B. J. (2005). Genetic relationships amongst macadamia varieties grown in South Africa as assessed by RAPD markers. *South Afr. J. Plant Soil* 22 (2), 71–75. doi: 10.1080/02571862.2005.10634684
- Peace, C. P., Vithanage, V., Neal, J., Turnbull, C. G. N., and Carroll, B. J. (2004). A comparison of molecular markers for genetic analysis of macadamia. *J. Hortic. Sci. Biotechnol.* 79 (6), 965–970. doi: 10.1080/14620316.2004.11511874
- Peace, C. P., Vithanage, V., Turnbull, C. G. N., and Carroll, B. J. (2002). Characterising macadamia germplasm with codominant radiolabelled DNA amplification fingerprinting (RAF) markers. *Acta Hort.* 575, 371–380. doi: 10.17660/ActaHortic.2002.575.42
- Peace, C. P., Vithanage, V., Turnbull, C. G. N., and Carroll, B. J. (2003). A genetic map of macadamia based on randomly amplified DNA fingerprinting (RAF) markers. *Euphytica* 134 (1), 17–26. doi: 10.1023/A:1026190529568

- Project, R. A. (2007). The rice annotation project database (RAP-DB): 2008 update. *Nucleic Acids Res.* 36 (suppl_1), D1028–D1033. doi: 10.1093/nar/gkm978
- Ranketse, M., Hefer, C. A., Pierneef, R., Fourie, G., and Myburg, A. A. (2022). Genetic diversity and population structure analysis reveals the unique genetic composition of South African selected macadamia accessions. *Tree Genet. Genomes* 18 (2), 15. doi: 10.1007/s11295-022-01543-0
- SAMAC (2020) Statistics of the Southern African macadamia industry. In: *Macadamias South Africa*. Available at: <https://www.samac.org.za/industry-statistics/> (Accessed 11/05 2020).
- Schmidt, A. L., Scott, L., and Lowe, A. J. (2006). Isolation and characterization of microsatellite loci from macadamia. *Mol. Ecol. Notes* 6 (4), 1060–1063. doi: 10.1111/j.1471-8286.2006.01434.x
- Sedgley, M. (1983). Pollen tube growth in macadamia. *Scientia Hort.* 18 (4), 333–341. doi: 10.1016/0304-4238(83)90015-8
- Sharma, P., Al-Dossary, O., Alsubaie, B., Al-Mssallem, I., Nath, O., Mitter, N., et al. (2021a). Improvements in the sequencing and assembly of plant genomes. *Gigabyte* 2021, 1–10. doi: 10.46471/gigabyte.24
- Sharma, P., Masouleh, A. K., Topp, B., Furtado, A., and Henry, R. J. (2021b). De novo chromosome level assembly of a plant genome from long read sequence data. *Plant J.* 109 (3), 727–736. doi: 10.1111/tpj.15583
- Sharma, P., Murigneux, V., Haimovitz, J., Nock, C. J., Tian, W., Kharabian Masouleh, A., et al. (2021c). The genome of the endangered *Macadamia janseni* displays little diversity but represents an important genetic resource for plant breeding. *Plant Direct* 5 (12), e364. doi: 10.1002/pld3.364
- 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 (3), 269–277. doi: 10.1023/A:1025025522276
- Storey, W. B., and Hamilton, R. A. (1953). “The macadamia nut industry in Hawaii,” in *California Avocado Society 1953-54 Yearbook* 38, 63–67. Available at: http://avocadosource.com/CAS_Yearbooks/CAS_38_1953-54/CAS_1953-54_PG_063-067.pdf
- Tang, Y., Yang, X., Cai, Y., Li, M., Zeng, L., Zheng, W., et al. (2018). Optimization and application of SSR-PCR reaction system for macadamia. *Fujian J. Agric. Sci.* 33 (2), 154–158. doi: 10.19303/j.issn.1008-0384.2018.02.009
- Tan, Q., Wang, W., Chen, H., Wei, Y., Zheng, S., Huang, X., et al. (2020). Genetic diversity analysis of macadamia varieties based on single-nucleotide polymorphism. *Mol. Plant Breed.* 18 (21), 7246–7253. doi: 10.13271/j.mpb.018.007246
- Tan, Q., Wang, W., Wei, Y., Zheng, S., Huang, X., He, X., et al. (2019). Diversity analysis of fruit traits related to yield in macadamia germplasm. *J. Fruit Sci.* 36 (12), 1630–1637. doi: 10.13925/j.cnki.gsx.20190087
- Tan, Q., Wei, Y., Huang, X., Zhang, T., Xu, P., Song, H., et al. (2021). Analysis of fruit characteristics and nutrients of 10 accessions of *Macadamia integrifolia*. *J. Fruit Sci.* 38 (5), 672–680. doi: 10.13925/j.cnki.gsx.20200372
- Toft, B. D., Alam, M., and Topp, B. (2018). Estimating genetic parameters of architectural and reproductive traits in young macadamia cultivars. *Tree Genet. Genomes* 14 (4), 50. doi: 10.1007/s11295-018-1265-x
- Topp, B. L., Nock, C. J., Hardner, C. M., Alam, M., and O’connor, K. M. (2019). “Macadamia (*Macadamia* spp.) breeding,” in *Advances in plant breeding strategies: Nut and beverage crops: Volume 4*. Eds. J. M. Al-Khayri, S. M. Jain and D. V. Johnson (Cham: Springer Nature Switzerland AG), 221–251. doi: 10.1007/978-3-030-23112-5_7
- Trueman, S. J. (2013). The reproductive biology of macadamia. *Scientia Hort.* 150, 354–359. doi: 10.1016/j.scienta.2012.11.032
- Urata, U. (1954). *Pollination requirements of macadamia* (Honolulu (HI: Technical Bulletin), 1–40.
- Vithanage, V., Hardner, C., Anderson, K. L., Meyers, N., Mcconchie, C., and Peace, C. (1998). Progress made with molecular markers for genetic improvement of macadamia. *Acta Hort.* 461, 199–208. doi: 10.17660/ActaHortic.1998.461.20H
- Vithanage, V., and Winks, C. W. (1992). Isozymes as genetic markers for macadamia. *Scientia Hort.* 49 (1), 103–115. doi: 10.1016/0304-4238(92)90147-5
- Yue, J., Liu, J., Tang, W., Wu, Y. Q., Tang, X., Li, W., et al. (2020). Kiwifruit genome database (KGD): a comprehensive resource for kiwifruit genomics. *Hortic. Res.* 7, 117. doi: 10.1038/s41438-020-0338-9
- Zeng, H., and Du, L. (2017). *Illustrated guide to identification of macadamia cultivars* (Beijing: China Agriculture Press).
- Zhang, M. (2011). *Study on the Fruit Main Ingredients of the Twenty-eight Macadamia Germplasm* (Hainan University, China: Master Thesis).
- Zhang, Z., Chai, M., Yang, Z., Yang, Z., and Fan, L. (2022). GRAND: an integrated genome, transcriptome resources, and gene network database for gossypium. *Front. Plant Sci.* 13, 773107. doi: 10.3389/fpls.2022.773107
- Zhang, T. T., He, H., Xu, C. J., Fu, Q., Tao, Y. B., Xu, R., et al. (2021). Overexpression of type 1 and 2 diacylglycerol acyltransferase genes (JcDGAT1 and JcDGAT2) enhances oil production in the woody perennial biofuel plant *Jatropha curcas*. *Plants (Basel)* 10 (4), 699. doi: 10.3390/plants10040699