Chromosomal-level genome of macadamia (*Macadamia integrifolia*)

Authors

Chengcai Xia, Sirong Jiang, Qiujin Tan, Wenquan Wang, Long Zhao, ..., Wenlin Wang[*], Meiling Zou[*], Zhiqiang Xia[*]
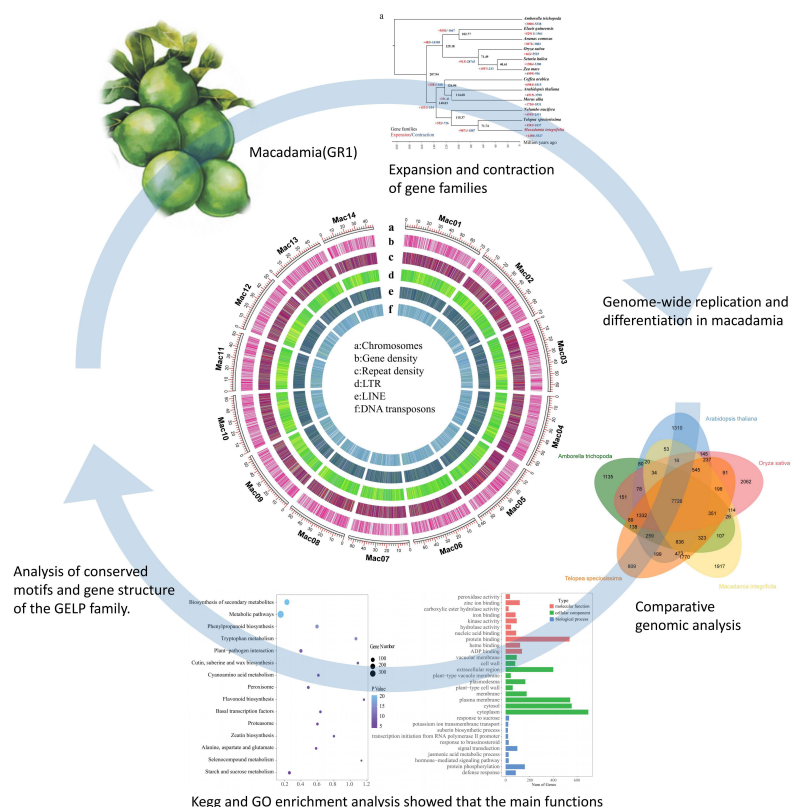
Correspondences

22312595@qq.com;
mlzou@hainanu.edu.cn;
zqiangx@gmail.com

In Brief

Macadamia is an important tropical commodity crop, and macadamia gene function research and evolutionary research require high-quality annotated gene sets and chromosome-level genomes. This study provides chromosome-level genome sequences to provide new resources for future macadamia omics research and molecular breeding.

Graphical abstract

Macadamia(GR1)

Expansion and contraction of gene families

a:Chromosomes
b:Gene density
c:Repeat density
d:LTR
e:LINE
f:DNA transposons

Genome-wide replication and differentiation in macadamia

Comparative genomic analysis

Kegg and GO enrichment analysis showed that the main functions

Analysis of conserved motifs and gene structure of the GELP family.

Highlights

- A chromosome-scale, high-quality macadamia GR1 reference genome was constructed through a combination of Nanopore sequencing, with 14 chromosomes

- Protein sequences of macadamia and 11 other species of the same family were compared to evaluate expansion and contraction in the macadamia gene family

- Proteaceae diverged from Nelumbonaceae nearly 115.37 million years ago and from Rubiaceae about 140 million years ago

- Identified 120 GELP family members using our assembled macadamia genome.

Chromosomal-level genome of macadamia (*Macadamia integrifolia*)

Chengcai Xia[1,2#], Sirong Jiang[1,2#], Qiujin Tan[3], Wenquan Wang[2], Long Zhao[4], Chenji Zhang[5], Yuting Bao[1,2], Qi Liu[1,2], Jianjia Xiao[1,2], Ke Deng[4], Miaohua He[1,2], Pengliang An[1,2], Wenlin Wang[3*], Meiling Zou[1,2*], and Zhiqiang Xia[1,2*]

[1] Sanya Nanfan Research Institute of Hainan University, Hainan Yazhou Bay Seed Laboratory, Sanya 572025, China
[2] College of Tropical Crops, Hainan University, Haikou 570288, China
[3] Guangxi South Subtropical Agricultural Science Research Institute, Longzhou 532415, China
[4] Academy of Agriculture and Forestry Sciences, Qinghai University, Xining 810016, China
[5] College of Agriculture, China Agricultural University, Beijing 100083, China
# These authors contributed equally: Chengcai Xia, Sirong Jiang
* Corresponding authors, E-mail: 22312595@qq.com; mlzou@hainanu.edu.cn; zqiangx@gmail.com

Abstract

Macadamia from the family Proteaceae is a plant native to Australia and has long been favoured by people for its crispy and high nutritional and medicinal value. Here, the genome of GUIRE 1 (GR1), a highly heterozygous superior cultivar of macadamia nut, was sequenced and assembled using nanopore sequencing, and a 807-Mb genome (contig N50, 1.9 Mb; scaffold N50, 54.70 Mb) and 14 chromosomes were obtained. A total of 453 Mb (about 55.95%) repetitive sequences and 37,657 protein-coding genes were obtained by gene annotation and homologous protein comparison. Proteaceae diverged from Nelumbonaceae nearly 115.37 million years ago and from Rubiaceae about 140 million years ago. A genome-wide duplication (WGD) event occurred in macadamia 41 million years ago based on the WGD analysis. The functional enrichment analysis of *M. integrifolia*-specific gene families revealed their roles in signal transduction, protein phosphorylation, protein binding, and defense response. Here, a highly heterozygous genome of *M. integrifolia* was unlocked to provide a database for breeding and molecular mechanism research.

INTRODUCTION

Macadamia, also known as Hawaii nut, belongs to the family Proteaceae, and *M. integrifolia* F. Mull., an evergreen tree without taproots and a shallow root system, is suitable for subtropical climate, and is classified as a subtropical fruit tree[1]. Macadamia nut contains more than 70% fat, 9% protein and eight essential amino acids; thus, its crispy kernels have abundant unsaturated fatty acids with high nutrition and health value[2–4]. Long-term consumption of macadamia nuts, which is known as the 'queen of dried fruits', contributes to the prevention and treatment of cardiovascular diseases[5,6]. In addition, the kernel, shell, and oil meal of macadamia are processed into various drinks, oil[7], activated carbon[8], adsorbent, feed, and other products. Furthermore, macadamia trees are beautifully shaped, with dense branches and foliage, pretty and fragrant flowers, and solid and compact wood, and are ideal insect-resistant landscaping plants.

Macadamia is endemic to the subtropical rainforests of eastern Australia and is one of the few crops to have been domesticated from dicotyledons[9]. The plant is concentrated in Australia, USA, Kenya, South Africa, Costa Rica, Guatemala, Brazil, and other tropical countries[10]. Macadamia is one of the few rapidly expanding crops in USA, Australia, South Africa, New Zealand, and China. Over 500 selected landraces and varieties are present. The macadamia germplasm in the world ranges from 34° N to 34° S and involves more than 20 countries, but most commercial production areas are located in 16°–24°

N, and the main production countries are China, Australia, South Africa, USA, and Kenya. In 2020, Macadamia had a global cultivation area of 6.05 million acres, with the largest cultivation area of 4.7 million acres in China. Thus, China has the highest macadamia production in the world.

Macadamia is highly heterozygous and predominantly outcrossing. Its genome size is 896 Mb, and all cultivars are diploid[11–14]. Study of its origin and evolutionary history requires a high-quality macadamia genome, and two macadamia genome research studies have been reported. The release of the HAES 741 macadamia genome opened a new era of macadamia genomics and was an important milestone in macadamia research. In 2016, Nock et al. used Illumina short-read sequence data to construct a highly fragmented genome sketch of HAES 741 with a total length of 518 Mb, 3,522-bp contig N50, and 4,745-bp scaffold N50, which annotated 35,337 protein-coding genes[15]. In 2020, they reassembled the HAES 741 genome using a combined strategy of Illumina short read sequences and PacBio long read sequences, with a 745-Mb genome and 413-kb scaffold N50. After fixing scaffolds to 14 chromosomes using seven genetic linkage maps, 34,274 protein-coding genes were predicted[12]. The studies of Ming et al.[16] and Nock et al.[12] used the world-famous macadamia nut variety Kau (*Macadamia integrifolia* Maiden & Betche) as the material and used three-generation sequencing to complete the genome assembly . The genome size is 794 Mb, contig N50 281 kb and they annotated 37,728 genes. They found the gene

families related to oil synthesis and fruit shell development in the macadamia genome have expanded. Then they selected 112 representative cultivated and wild macadamia materials from Hawaii and Australia to resequence, and the results revealed a phylogenetic relationship of macadamia populations. Through 2–3 generations of artificial selection, selective clearance areas are generated, which provides genomic basis and direct evidence for the 'one-step' theory of asexual crop domestication. These findings provide a theoretical basis for the rapid domestication of new species[16]. At present, two studies on the macadamia genome have been reported, and both are for the same species. It is also the only genome in macadamia, falling short of the genome resources required to study the plant.

GDSL esterase/lipase protein (GELP) is a hydrolase that hydrolyzes thioesters, aryl esters, phospholipids, amino acids, and other substrates. This type of lipase is widely present in prokaryotes and eukaryotes. At present, GELP is widely involved in physiological activities such as normal plant growth and development, organ morphogenesis, secondary metabolism and stress, and plays an important role in oil metabolism of oil crop seeds[17,18]. Although GELP family genes have been identified and studied in other plants, they have not been identified in macadamia.

GUIRE 1 (GR1), an excellent cultivar bred in China, has evident advantages such as early fruit setting, high and stable yield, high quality, and strong stress resistance. The high-quality GR1 assembled genome described in this study will help in seed selection and breeding of macadamia and accelerate research on the molecular mechanisms in macadamia, thereby providing a database for macadamia researchers.

RESULTS

Genome assembly

The genome of GR1, a highly heterozygous cultivar, was assembled from the beginning using 281 Gb long-read and long-sequence data obtained by nanopore sequencing. The total genome size was about 924 Mb, including 1,757 contigs, and the contig N50 and GC content were 1.97 Mb and 39.28%, respectively. Busco's evaluation of genome assembly quality and integrity showed that 95.7% of plant single-copy homologous genes were complete, and 74.3% of complete single-copy and 21.4% of multi-copy genes were complete. The preliminarily assembled macadamia genome had high quality and coverage. Using the macadamia genome published by Nock et al.[12] as the reference genome, the initially assembled GR1 genome was constructed to the chromosomal level, resulting in a super high-density genome map with a 792-Mb genome size, 46.24-Mb scaffold N50, and 14 chromosomes (Fig. 1; Table 1). The size of the GR1 genome we finally assembled is consistent with that of Kau (Supplemental Table 1). BUSCO was used to assess the quality and integrity of genome assembly, and 89.7% of plant single-copy homologous genes were intact. Complete single- and multi-copy genes accounted for 73% and 16.7%, respectively.

Genome annotation

The repetitive sequences of the GR1 genome were obtained by homologous and de novo annotations to understand the

Table 1. GR1 genome assembly statistics.

Statistic	Contig length (bp)	Contig no.	Scaffold length (bp)	Scaffold no.
N50	1,971,883	124	55,868,876	7
N80	494,409	388	50,524,966	11
N90	166,444	729	47,725,559	13
Longest	30,625,757	–	77,511,179	–
Total	924,331,529	1,757	792,388,653	14

genomic characteristics of GR1. About 56.77% of the genome was identified as duplicate regions. Among these regions, long terminal repeats (189 Mb, 23.96% of the genome) were identified as the main repeat sequences, followed by long interspersed nuclear elements (44 Mb, 5.63% of the genome) (Supplemental Table 2). A total of 2,543 tRNAs, 1,119 rRNAs, 248 microRNAs, and 1,562 snRNAs as non-coding genes were present in the GR1 genome. MAKER software combined with the AUGUSTUS pipeline ab initio gene prediction was used for protein-coding gene prediction, and proteins were compared to perform gene functional annotation for GR1. Finally, 37,657 protein-coding genes were obtained, which are consistent with the number of protein coding genes contained in 'Kau'. (Supplemental Table 3).

Expansion and contraction of gene families

Protein sequences of macadamia and 11 other species of the same family were compared to evaluate expansion and contraction in the macadamia gene family. These species were *Telopea speciosissima* (Proteaceae), dicotyledons: *Arabidopsis thaliana*, *Coffea arabica*, and *Morus alba*; monocotyledons: *Setaria italica*, *Oryza sativa*, *Zea mays*, *Elaeis guineensis*, *Ananas comosus*, and *Nelumbo nucifera* from the same family (Proteaceae) were used for gene family clustering, with *Amborella trichopoda* as the outgroup. OrthoVenn2 (https://orthovenn2.bioinfotoolkits.net/home) was used to identify gene families unique to all species. In macadamia, 37,657 genes were divided into 14,930 gene families, and 13,613 single-copy gene families were present.

A phylogenetic tree was constructed using the single-copy genes of these single-copy gene families of the 12 species, with *Amborella trichopoda* as the outgroup. Evolutionarily, macadamia diverged from waratah, a member of the same family about 71.74 million years ago. Proteaceae diverged from Nelumbonaceae nearly 115.37 million years ago. Proteaceae diverged from Rubiaceae about 140 million years ago (Fig. 2a). Expansion or contraction of the gene family is an important feature in selective evolution. Compared with *Nelumbo nucifera* of the same family, macadamia evolution involved new genes and gene families, but their evolution occurred independently with a different degree of gene families being lost for each species. In macadamia, 1,498 gene families experienced expansion, and 5,327 gene families experienced contraction (Fig. 2a).

Genome-wide replication and differentiation in macadamia

On the basis of the expansion and contraction of gene families, the synonymous substitution rates (Ks) of the homologous genes of macadamia, waratah, and *A. thaliana* were calculated, and Ks of the aligned homologous genes of different species was calculated to obtain the Ks curve of the genome and infer the WGD[19] event of macadamia. An evident
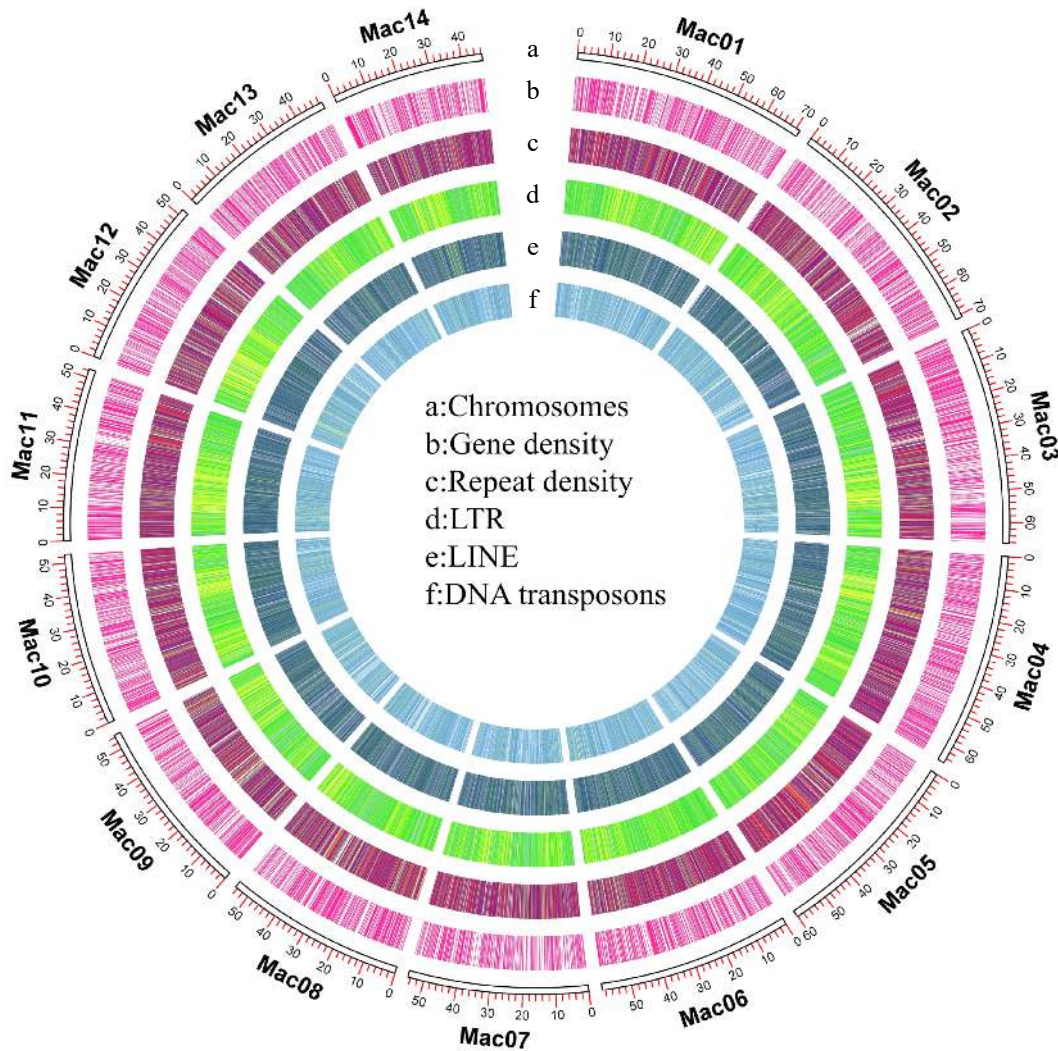
Fig. 1 Overview of the macadamia genome. (a) Chromosomes, (b) gene density, (c) repeat density, (d) long terminal repeats (LTRs), (e) long interspersed nuclear elements (LINEs), (f) DNA transposons.

peak between macadamia and waratah at Ks value of 0.34 indicated that a WGD event occurred in macadamia about 41 million years ago, and this WGD event might have been a doubling event shared by the Proteales (Fig. 2b).

Comparative genomic analysis

Comparative genomic analysis was performed using the genomes of macadamia, *T. speciosissima*, *A. thaliana*, *A. trichopoda*, and *O. sativa*. The results showed 7,726 gene families in the five species. The macadamia, *T. speciosissima, A. thaliana, A. trichopoda,* and *O. sativa* gene families were 1,917, 809, 1,310, 1,135 and 2,062, respectively (Supplemental Fig. 1). The Kyoto Encyclopedia of Genes and Genomes (KEGG)[20] analysis of the specific gene families in macadamia showed that these gene families were predominantly concentrated in metabolism-related biological functions such as metabolic pathways, biosynthesis of secondary metabolites, phenylpropanoid biosynthesis, and tryptophan metabolism (Fig. 3a). Further, the Gene Ontology (GO) enrichment analysis showed that the main functions of these gene families were signal transduction, protein phosphorylation, protein binding, and defense response (Fig. 3b).

Genome-wide identification of GELP genes in *M. integrifolia*

We identified 120 GELP family members unevenly distributed on 14 chromosomes in macadamia, and most genes were located on chromosome 10 (Supplemental Fig. 2). In gene structure analysis, some differences were observed in the exon and intron structures of GELP family members. The number of exons of GELP family members ranged from 2 to 16, and the number of introns ranged from 1 to 15 (Fig. 4). Motif analysis revealed high motif conservation among GELP family members. Most GELP family genes contained motif1 and motif4, indicating that these two conserved motifs are particularly important in the GELP family. CAAT and TATA boxes were the most *cis*-acting elements in the promoters of all GELP family members, and they were present in all members of the GELP gene family. Among the GELP family members, 280 gene promoters contained *cis*-acting elements necessary for anaerobic induction, and 233 gene promoters contained abscisic acid *cis*-acting elements, 230 gene promoters contained methyl jasmonate *cis*-acting elements, 101 gene promoters contained salicylic acid *cis*-acting elements, 38 gene promoters contained auxin *cis*-acting elements, and 52 gene promoters contained gibberellin
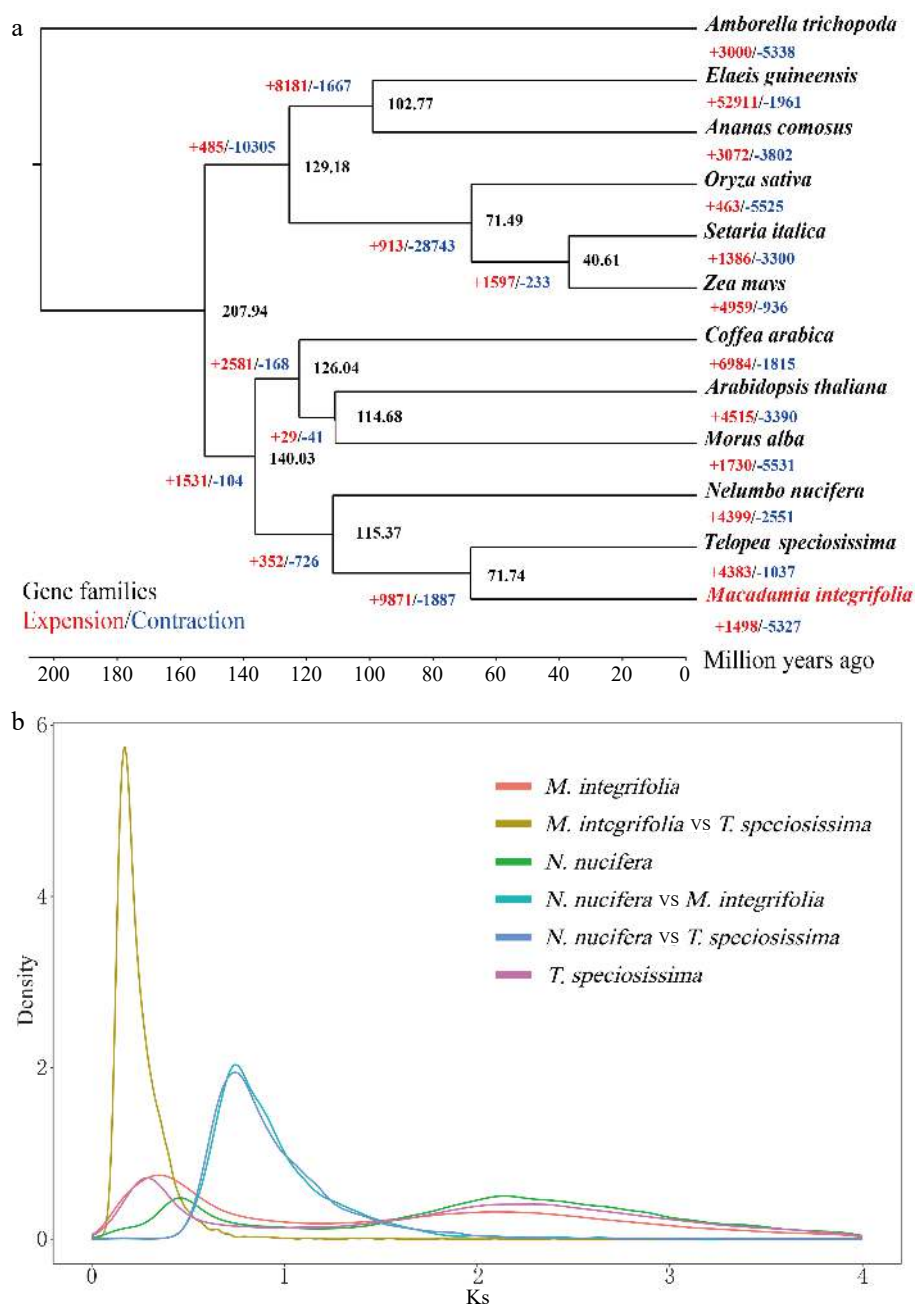
Fig. 2 Genome evolution analysis. (a) Phylogenetic tree of 12 species constructed using single-copy orthologs. (b) Frequency distribution of synonymous substitution rates (Ks) between homologous gene pairs in the syntenic blocks of *M. integrifolia*, *M. integrifolia* vs *T. speciosissima*, *N. nucifera*, *N. nucifera* vs *M. integrifolia*, *N. nucifera* vs *T. speciosissima*, and *T. speciosissima*.

cis-acting elements. In addition, there were *cis*-acting elements for circadian control and defense and stress response (Supplemental Fig. 3).

DISCUSSION

Macadamia has important economic value in the global food industry. A high-quality genome sequence is needed as a basis to promote research on the molecular mechanisms of macadamia and accelerate the breeding of superior varieties. GR1 is the first macadamia variety in China that is protected by intellectual property rights. This variety features early fruit setting, high yield, and strong stress resistance. The macadamia

genome is relatively complex with high heterozygosity and outcrossing. To elucidate the genetic system and evolution of Proteaceae, we sequenced the genome of macadamia, providing new insights and genomic resources for breeding. We report a high-quality chromosome-scale genome assembly of passion fruit, with a contig N50 of 1.9 Mb and assembly to 14 pseudo-chromosomes. This reference genome is higher in continuity than the previously published macadamia genome, such as HAES 741, with a scaffold N50 ~413Kb. The high quality of our assembly can be attributed to the use of the unique combination of Nanopore sequencing[21,22] with chromosome-scale scaffolding via RagTag. The macadamia genome sequence provides an important resource for future molecular
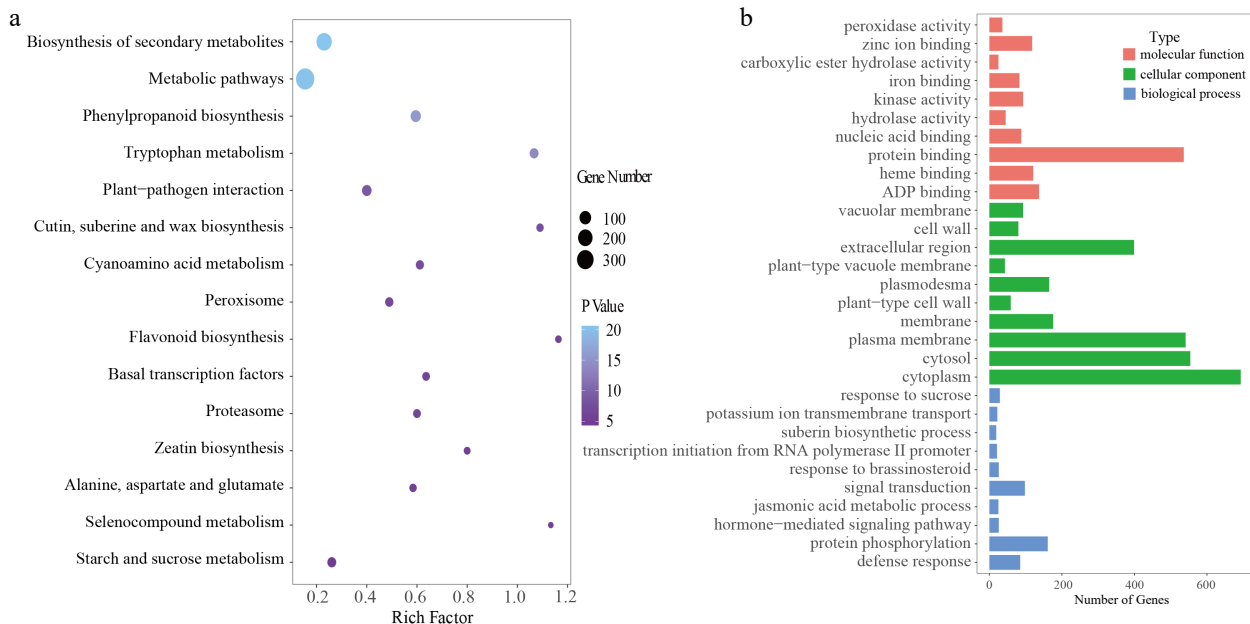
Fig. 3 Gene family analysis. (a) KEGG analysis of unigenes in the GR1 genome, (b) GO analysis of unigenes in the GR1 genome.

breeding and evolutionary studies.

This annotated chromosome-level reference genome of macadamia can provide important information on the gene content, duplication elements, gene location on 14 chromosomes and RNA types in macadamia. In addition, Macadamia is the most important genus of the family Proteaceae, and this high-quality assembled macadamia genome revealed the important role of macadamia in evolutionary history. The divergence between macadamia and its relative *T. speciosissima* occurred about 71.74 million years ago. The WGD event, a paleopolyploidy event, is common in plants, and it indicates the development of new gene functions or the formation of a new species. About 115.37 million years ago, Proteaceae and Nelumbonaceae differentiated to form a separate family, which provided a data base for future research on the evolutionary relationship of Proteales. This genomic information for macadamia will help clarify the evolutionary processes in Proteaceae species and contribute to improving the understanding of the physiological and morphological diversity of Proteaceae species.

Macadamia is rich in oil and unsaturated fatty acids, and the GELP gene family is involved in the regulation of plant growth and development, secondary metabolism and oil synthesis in fruits[23–26]. However, previous studies have found that GELP gene family members have been identified in *Arabidopsis*[27], rice, rape[24] and other plants, but studies of this gene family have not been reported in macadamia. To aid in future studies on oil synthesis in macadamia fruit as well as other molecular mechanisms, we sought to identify 120 members using our assembled macadamia gene genome. The number of GELP family members varies among species, which may be because of the different degree of evolution of the GELP family. Chromosome mapping analysis showed that GELP family genes were unevenly distributed on 14 chromosomes of macadamia. Gene structure analysis showed that most GELP genes contained 5 exons and 4 introns. GELP is structurally conservative. Promoter *cis*-acting element analysis showed that the pro-

moters of GELP family members contained *cis*-acting elements related to hormone response and plant growth and metabolism. Therefore, the functions of GELP genes may be related to hormone response and growth and metabolism of plants.

MATERIALS AND METHODS

Materials and sequencing

Fresh leaves were collected from the Macadamia Germplasm Resource Nursery of Guangxi Institute of South Subtropical Agricultural Sciences, frozen in the field, and stored at −80 °C until DNA extraction. High-molecular weight genomic DNA was extracted from fresh-frozen macadamia leaves by the modified CTAB[28] method. Nanopore sequencing was adopted to obtain the long-read sequence data of 281 Gb.

Chromosome-level genome assembly

The chromosome-level genome assembled by Nock et al.[12] was downloaded from the NCBI (www.ncbi.nlm.nih.gov/genome/?term=Macadamia+integrifolia) database for use as the reference genome, and the genome of the Contig I version of GR1 was intended to be scaled to the chromosome level. The genomic size of cultivar 'HAES 741' assembled by Nock et al. was 744.64 Mb, with a scaffold N50 of 413.4 kb[12]. First, Nanopore sequence data were assembled *de novo* using NextDenovo. Redundans were used to remove the genome duplication[29]. Second, sequence errors from the preliminary assembly of the genome were removed using NextPolish[30]. Third, non-chromosomal sequences were removed from the downloaded reference genome sequences and only the chromosomal sequences were retained. Fourth, RagTag tools[31] were used for scaling the assembled GR1 genome to the chromosome level, and the GR1 genome sequence at the chromosome level was constructed. Finally, the integrity of the genome assembly based on a single-copy homologous plant-specific database was assessed using the benchmarking universal single-copy orthologs (BUSCO)[32] default setting.

Fig. 4 Analysis of conserved motifs and gene structure of the GELP family.

Gene prediction and functional annotation

Genome repeats were annotated from scratch by the RepeatMasker program (www.repeatmasker.org). Protein-coding genes were predicted by *ab initio* calculations, conserved protein homologs, and combinations of assembled transcripts based on repeated masked genomes. The homology prediction of protein sequences was performed using the macadamia genome HAES 741[12], waratah genome of family Proteaceae, and GR1 genome. In addition, transcriptome data of macadamia in different tissues at different periods was downloaded from the NCBI database. Transcriptome data were assembled by Trinity and used for preliminary annotation. The Augustus software (http://bioinf.uni-greifswald.de/augustus) was used for gene prediction and annotation[33]. The MAKER[34]

software was used to integrate these results in the final genetic model.

Phylogenetic tree construction and evolution rate estimation

The OrthoMCL[35] program was used to identify macadamia, waratah, lotus flower, and the clusters of homologous genes of eight other species (*A. thaliana*, *C. arabica*, *M. alba*, *S. italica*, *O. sativa*, *Z. mays*, *E. guineensis*, and *A. comosus*) using the FastTree algorithm (v2.1.9)[36]. The maximum likelihood (ML) tree of single-copy homologous genes was constructed. This ML tree was converted to a super-time-scale phylogenetic tree by r8s[37] using the calibration time of the TimeTree website.

Syntenic analysis

The MCScanX toolkit[38] with default parameters was used to identify collinear blocks. Protein names were used as search queries for genomes of other plant species to find the best matching pair. Each aligned block represented homologous pairs derived from a common ancestor. The Nei–Gojobori method implemented in phylogenetic analysis by maximum likelihood (PAML)[39] was adopted to calculate the synonymous substitution rates (Ks) of homologs in the collinear region. The mean Ks value was considered to represent the collinear region.

Gene family analysis and comparison

The number of gene families in macadamia, *T. speciosissima*, *A. thaliana*, *O. sativa*, and *A. trichopoda* was calculated using the OrthoVenn2 website (https://orthovenn2.bioinfotoolkits.net/home). Besides macadamia-specific gene families, the protein sequences of specific gene families were screened and submitted to the KOBAS online website for GO and KEGG enrichment analyses.

Pfam (http://pfam.xfam.org/) was used to obtain the characteristic domain of the GELP gene family protein[40]. To build a more accurate hidden Markov model and predict all GELP family members in macadamia, the GELP family members in the GR1 genome were searched and screened by Hmmer 3.0, the screened sequences were compared with ClustalW, and the hidden Markov model of these verified sequences was constructed by hmmbuild. Finally, 120 GELP family members were screened. The gene structure of GELP family members was analyzed by TBtools[41] and the macadamia annotation file. The GELP family protein motifs were analyzed using MEME (http://meme-suite.org) and visualized by TBtools. A 2000-bp sequence upstream of the GELP start codon (ATG) was obtained from the macadamia genome database as the promoter sequence. The *cis*-acting elements of promoters were predicted using the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) and visualized by TBtools.

Data availability

The whole genome sequence data reported in this paper have been deposited in the Genome Warehouse in China National Center for Bioinformation, under accession number PRJCA008938 that is publicly accessible at https://ngdc.cncb.ac.cn/gwh.

ACKNOWLEDGMENTS

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. Tan Q, Wang W, Wei YR, Zheng SF, Huang XY, et al. 2019. Diversity analysis of fruit traits related to yield in Macadamia germplasms. *Journal of Fruit Science* 36:1630−37
2. Wall MM. 2010. Functional lipid characteristics, oxidative stability, and antioxidant activity of macadamia nut (*Macadamia integrifolia*) cultivars. *Food Chemistry* 121:1103−8
3. Birch J, Yap K, Silcock P. 2010. Compositional analysis and roasting behaviour of gevuina and macadamia nuts. *International Journal of Food Science and Technology* 45:81−86
4. Maguire LS, O'Sullivan SM, Galvin K, O'Connor TP, O'Brien NM. 2004. Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. *International Journal of Food Sciences and Nutrition* 55:171−78
5. Garg ML, Blake RJ, Wills RBH. 2003. Macadamia nut consumption lowers plasma total and LDL cholesterol levels in hypercholesterolemic men. *The Journal of Nutrition* 133:1060−63
6. Garg ML, Blake RJ, Wills RBH, Clayton EH. 2007. Macadamia nut consumption modulates favourably risk factors for coronary artery disease in hypercholesterolemic subjects. *Lipids* 42:583−87
7. Liu J, Huang L. 2005. The nutritional value of macadamia and its development and utilization. *Food and Nutrition in China* 2:25−26
8. Tu X, Zhang X, Liu Y, Du L, Huang M, et al. 2015. Study on the technology of activated carbon pre paration of microwave irradiation of macadamia shell. *Science and Technology of Food Industry* 36:253−59
9. Geng J, Tao L, Yue H, Li Z, He X. 2021. Review on Comprehensive Utilization of Macadamia Nutshell. *Tropical Agricultural Science & Technology* 38:41−47
10. Nock CJ, Hardner CM, Montenegro JD, Termizi AAA, Batley J. 2019. Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. *Frontiers in Plant Science* 10:334
11. Topp BL, Nock CJ, Hardner CM, Alam M, O'Connor, et KM. 2019. Macadamia (Macadamia spp.) breeding. In *Advances in Plant Breeding Strategies: Nut and Beverage Crops*, eds. Al-Khayri J, Jain S, Johnson D. Switzerland: Springer International Publishing. pp. 221−51 https://doi.org/10.1007/978-3-030-23112-5_7
12. Nock CJ, Baten A, Mauleon R, Langdon KS, Topp B, et al. 2020. Chromosome-scale assembly and annotation of the Macadamia genome (*Macadamia integrifolia* HAES 741). *G3 Genes|Genomes|Genetics* 10:3497−3504
13. Stace HM, Douglas AW, Sampson JF. 1998. Did 'Paleo-polyploidy' Really occur in Proteaceae? *Australian Systematic Botany* 11:613−29
14. Nock CJ, Elphinstone MS, Ablett G, Kawamata A, Hancock W, et al. 2014. Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild *Macadamia* (Proteaceae). *Applications in Plant Sciences* 2:1300089
15. Nock CJ, Baten A, Barklaet BJ, Furtadoal A, Henry RJ, et al. 2016. Genome and transcriptome sequencing characterises the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics* 17:937

16. Lin J, Zhang W, Zhang X, Ma X, Zhang S, et al. 2022. Signatures of selection in recently domesticated macadamia. *Nature Communications* 13:242

17. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF. 2004. GDSL family of serine esterases/lipases. *Progress in Lipid Research* 43:534−52

18. Chepyshko H, Lai CP, Huang LM, Huang LM, Liu JH, Shaw JF. 2012. Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (*Oryza sativa* L. *japonica*) genome: New insights from bioinformatics analysis. *BMC Genomics* 13:309−27

19. Otto SP. 2007. The evolutionary consequences of polyploidy. *Cell* 131:452−62

20. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 27:29−34

21. Belser C, Istace B, Denis E, Dubarry M, Baurens FC, et al. 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants* 4:879−87

22. Jain M, Koren S, Miga KH, Quick J, Rand AC, et al. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* 36:338−45

23. Takahashi K, Shimada T, Kondo M, Tamai A, Mori M, et al. 2009. Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in *Arabidopsis thaliana*. *Plant and Cell Physiology* 51:123−31

24. Dong X, Yi H, Han CT, Nou IS, Hur Y. 2016. GDSL esterase/lipase genes in *Brassica rapa* L.: genome-wide identification and expression analysis. *Molecular Genetics and Genomics* 291:531−42

25. Kim GK, Kwon SJ, Jang YJ, Chung JH, Nam MH, et al. 2014. GDSL lipase 1 regulates ethylene signaling and ethylene-associated systemic immunity in *Arabidopsis*. *FEBS Letters* 588:1652−58

26. Ding L, Guo X, Li M, Fu Z, Yan S, et al. 2018. Improving seed germination and oil contents by regulating the *GDSL* transcriptional level in *Brassica napus*. *Plant Cell Reports* 38:243−53

27. Ling H. 2008. Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana*. *Pakistan Journal of Biological Sciences* 11:763−67

28. Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15:8−15

29. Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research* 44:e113

30. Hu J, Fan J, Sun Z, Liu S. 2019. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36:2253−55

31. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, et al. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology* 20:224

32. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210−12

33. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, ea tl. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research* 34:W435−W439

34. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, ea tl. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18:188−96

35. Li L, Stoeckert SC Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13:2178−89

36. Price MN, Dehal PS, Arkin A. 2010. PFastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490

37. Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution* 7:1812−19

38. Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40:e49

39. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586−91

40. Ni P, Ji X, Guo D. 2020. Genome-wide identification, characterization, and expression analysis of GDSL-type esterases/lipases gene family in relation to grape berry ripening. *Scientia Horticulturae* 264:109162

41. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 1194−202