Fu et al. *Horticulture Research* (2021)8:35
https://doi.org/10.1038/s41438-021-00487-1

Horticulture Research
www.nature.com/hortres

ARTICLE Open Access

Combined genomic, transcriptomic, and metabolomic analyses provide insights into chayote (*Sechium edule*) evolution and fruit development

Anzhen Fu[1,2], Qing Wang[1], Jianlou Mu[2], Lili Ma[1,2], Changlong Wen[1], Xiaoyan Zhao[1], Lipu Gao[1], Jian Li[3], Kai Shi[4], Yunxiang Wang[5], Xuechuan Zhang[6], Xuewen Zhang[6], Fengling Wang[6], Donald Grierson[7] and Jinhua Zuo[1]

Abstract

Chayote (*Sechium edule*) is an agricultural crop in the Cucurbitaceae family that is rich in bioactive components. To enhance genetic research on chayote, we used Nanopore third-generation sequencing combined with Hi–C data to assemble a draft chayote genome. A chromosome-level assembly anchored on 14 chromosomes (N50 contig and scaffold sizes of 8.40 and 46.56 Mb, respectively) estimated the genome size as 606.42 Mb, which is large for the Cucurbitaceae, with 65.94% (401.08 Mb) of the genome comprising repetitive sequences; 28,237 protein-coding genes were predicted. Comparative genome analysis indicated that chayote and snake gourd diverged from sponge gourd and that a whole-genome duplication (WGD) event occurred in chayote at 25 ± 4 Mya. Transcriptional and metabolic analysis revealed genes involved in fruit texture, pigment, flavor, flavonoids, antioxidants, and plant hormones during chayote fruit development. The analysis of the genome, transcriptome, and metabolome provides insights into chayote evolution and lays the groundwork for future research on fruit and tuber development and genetic improvements in chayote.

Introduction

Chayote (*Sechium edule*) is a diploid perennial herbaceous climbing plant with 28 chromosomes ($2n = 2x = 28$)[1] that belongs to the Cucurbitaceae family[2]. The chayote fruit is a gourd that is consumed as a vegetable, also called vegetable pear, chuchu, pear squash, or sayote[3,4], and the tubers are also utilized as food source. Chayote is believed to have originated in Mexico, where it was first cultivated approximately 500 years ago[5,6]. It is commonly cultivated in tropical and subtropical areas, such as Brazil, India, Costa Rica, China, and Mexico, and is a significant commercial crop worldwide[7,8]. Chayote contains abundant bioactive compounds, such as phenolics, flavonoids, carotenoids, and bioactive polysaccharides[9–11] in the fruit, leaves, tubers, and stems[4,12], and has potential for the treatment of hypertension, diabetes, and inflammation, as well as other pharmacological applications[2,13,14].

In recent years, reports about chayote have been mainly related to the function, compositions, and applications of its fruit, stems, leaves, and tubers[4,9,15,16]. Although there has been extensive exploration of the genomes within the Cucurbitaceae family, such as

Correspondence: Donald Grierson (Donald.Grierson@nottingham.ac.uk) or Jinhua Zuo (zuojinhua@126.com)
[1]Key Laboratory of Vegetable Postharvest Processing, Ministry of Agriculture, Beijing Key Laboratory of Fruits and Vegetable Storage and Processing, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China) of Ministry of Agriculture, Key Laboratory of Urban Agriculture (North) of Ministry of Agriculture, The Collaborative Innovation Center of Cucurbits Crops, Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[2]College of Food Science and Technology, Hebei Agricultural University, Baoding 071001, China
Full list of author information is available at the end of the article
These authors contributed equally: Anzhen Fu, Qing Wang, Jianlou Mu

© The Author(s) 2021
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Cucumis sativus ($2n = 2x = 14$, 226.2 Mb)[17], *Cucumis melo* ($2n = 2x = 24$, 398.57 Mb)[18], *Citrullus lanatus* ($2n = 2x = 22$, 353.5 Mb)[19], *Cucurbita pepo* ($2n = 2x = 20$, 263 Mb)[20], *Luffa cylindrica* ($2n = 2x = 26$, 656.19 Mb)[21], *Lagenaria siceraria* ($2n = 2x = 22$, 313.4 Mb)[22], and *Benincasa hispida* ($2n = 2x = 24$, 913 Mb)[23], there is no complete genome of chayote available at present[24], and there have only been a few studies on methods for chayote DNA extraction[25] and on the genes expressed in chayote fruit[26,27].

Genome sequences provide resources for studies on evolution, genetic variation and traits for crop improvement[28]. To advance research on chayote, we assembled a draft genome using Nanopore and Hi–C data. Repeat sequences were identified, the functions of protein-coding genes were annotated, and expanded gene families were identified. Comparative genomics analysis indicated that of the species we studied, chayote is most closely related to snake gourd. Evidence for a paleoploidization (whole-genome duplication (WGD)) event in chayote was detected. The genes and mechanisms involved in fruit texture, pigment, flavor, plant hormones, and antioxidant properties were investigated. The chromosome-scale assembly allows a better understanding of evolution in Cucurbitaceae and provides insights for trait modification and breeding.

Results

Genome sequencing, assembly, and evaluation

Two libraries were built for the Illumina sequencing platform, and 39.01 Gb of high-quality data were obtained after filtration, with a total sequencing depth of approximately 55 with Q20 and Q30 percentages of 97.09 and 92.02, respectively. From the total kmer number and average kmer depth, the chayote genome length was estimated to be 710.23 Mb (Fig. S1). The proportion of repeat sequences was estimated to be approximately 62.93%, the heterozygosity rate was approximately 0.03%, and the GC content of the genome was approximately 40% (Fig. 1).

Approximately, 100.56 Gb of raw data was obtained. After data quality control (QC), the clean data volume was 91.97 Gb, representing a 151× sequencing depth. The clean data contained 4,155,091 reads with a read N50 of 29.68 kb and average read length of 22.13 kb (Table 1). Finally, a total of 608.17 Mb of genome sequence in 356 contigs (N50 of 10.09 Mb) with 38.71% GC content was obtained (Table 1). A total of 99.4% clean reads were mapped to the Nanopore reference genome. The CEGMA v2.5 database[29] contains 458 conserved core genes and 248 highly conserved genes from eukaryotes. The chayote assembled genome contained 445 core eukaryotic genes (CEGs) (97.16%) and 223 highly conserved CEGs (89.92%). Benchmarking universal single-copy orthologs (BUSCO) v2.0 software[30] was used to assess the integrity of the genome assembly, covering 2121 conserved core genes, and 2028 (95.62%) complete BUSCOs were found, which contained 1780 (83.92%) single-copy, 248 (11.69%) duplicated, 23 (1.08%) fragmented genes and 70 (3.30%) missing BUSCOs (Table 1).

Chromosome-level assembly of Hi–C

Hi–C technology is derived from chromosome conformation capture technology combined with high-throughput sequencing. It evaluates the interactions of chromosomes in three-dimensional space by capturing and sequencing the interactions between DNA fragments in chromosomes, information that can contribute to genome assembly. Ultimately, 606.42 Mb of genome sequence was anchored on 14 chromosomes by Hi–C assembly and manual adjustment, accounting for 97.04% of all genome sequence (Table S1). After correction of chromosome order and direction, a genome sequence

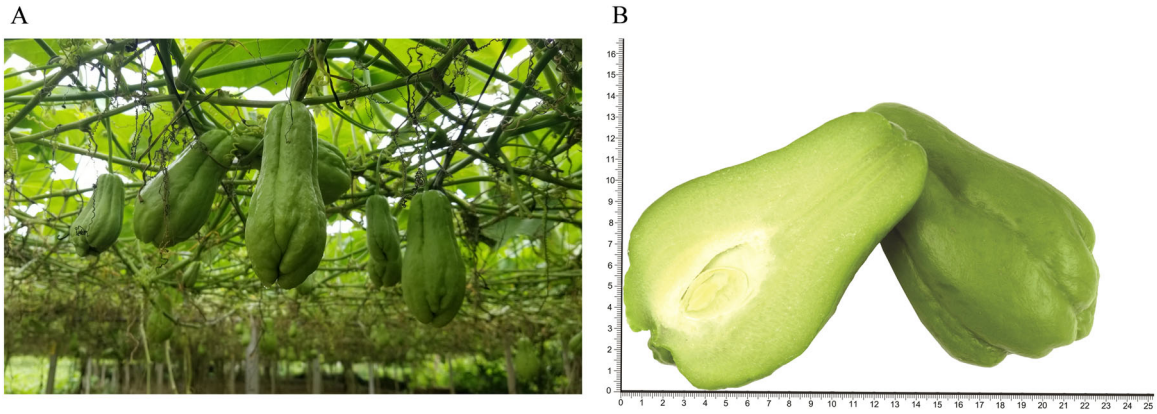

Fig. 1 Chayote fruits. A Fruiting chayote plants grown in Jianshui, Yunnan Province. The fruit photographed was approximately 10 × 18 cm; **B** longitudinal section of mature chayote fruits on the market

Table 1 Statistics of genome sequencing, Hi–C assembly and gene assessment in BUSCO

Parameter	Value
Nanopore sequencing	
Contig number	356
Contig N50 (Mb)	10.09
Genome size (Mb)	608.17
Hi–C assembly	
Scaffold number	103
Scaffold N50 (Mb)	46.56
Contig number	473
Contig N50 (Mb)	8.40
Genome size (Mb)	606.42
BUSCO evaluation	
Complete BUSCOs	2028 (95.62%)
Complete and single-copy BUSCOs	1780 (83.92%)
Complete and duplicated BUSCOs	248 (11.69%)
Fragmented BUSCOs	23 (1.08%)
Missing BUSCOs	70 (3.30%)
Total lineage BUSCOs	2121

length of 598.48 Mb was obtained, accounting for 98.69% of the total sequence length, with contig N50 and scaffold N50 of 8.40 and 46.56 Mb, respectively (Tables 1 and S2). The Hi–C heat map (Fig. 2A) shows the effectiveness of the genome assembly; an overview of the genome assembly is shown in Fig. 2B.

Genome annotation analysis

In total, 401.08 Mb of repetitive sequences were identified by analysis of the genome database and structure predictions, representing 65.94% of the whole genome. Long terminal repeats (LTRs) accounted for the highest proportion (36.96%) of the genome, including 9.11% Copia-LTRs and 27.85% Gypsy-LTRs (Table S3). Ab initio predictions, homology-based predictions and RNA-seq fragments were used to predict the genetic structure (Fig. S2), and 28,237 protein-coding genes were detected (Table S4). The average numbers of exons and introns per gene were 5.70 and 4.70, respectively (Table S5). The noncoding RNAs identified included 101 microRNAs, 1873 tRNAs and 298 rRNAs. Through BLA comparison and GeneWise, 1085 pseudogenes were found. Kyoto encyclopedia of genes and genomes (KEGG), KOG (eukaryotic orthologous groups), and gene ontology (GO) were used for functional annotation analysis and produced results for 97.28% of the assembled genome, including GO (53.19%), KEGG (32.31%), KOG (52.58%), TrEMBL (97.16%), and Nr (97.22%) (Table 2). Only 2.72% of the genome sequence was unannotated. Comparisons of repetitive sequence percentages and protein-coding gene numbers in chayote and ten other Cucurbitaceae are summarized in Fig. 3.

Comparative genomics analysis

To study the evolution of the chayote genome, we conducted comparative genomics analysis of 14 species, including 11 cucurbits (*S. edule*, *C. sativus*, *C. lanatus*, *L. cylindrica*, *L. siceraria*, *Cucurbita moschata*, *Trichosanthes anguina*, *Momordica charantia*, *C. pepo*, *B. hispida*, *C. melo*) and *Arabidopsis thaliana*, *Vitis vinifera*, and *Amborella trichopoda*. A total of 35,291 gene families were analyzed, representing 2,854 common gene families, and there were 134 specific gene families in the chayote genome (Fig. S3). These orthogroup gene statistics were calculated for each species, and chayote contained 36.3% single-copy genes and 36.4% two-copy genes (Fig. 4A). Cluster analysis of gene families was performed for *S. edule, T. anguina, L. cylindrica, C. lanatus* and *L. siceraria* (Fig. 4B), and GO and KEGG enrichment were used to analyze the specific gene families in chayote (*S. edule)* (Fig. S4). The unique genes were related to binding, catalytic activity, and metabolic and cellular processes. KEGG analysis showed that they were enriched in mismatch repair, DNA replication, homologous recombination, and nucleotide excision repair.

We constructed an evolutionary tree from 832 single-copy protein sequences (Fig. 4C) and found that *S. edule* and *T. anguina* (snake gourd) had the closest evolutionary relationship (27–45 Mya) and were estimated to have separated from *L. cylindrica* at 29–55 Mya. We predicted 184 contracted (Fig. S5) and 200 expanded (Fig. S6) gene families (Fig. 4C), which, according to GO and KEGG enrichment annotations, were enriched for glycosaminoglycan degradation, stilbenoid, diarylheptanoid, flavonoid and gingerol biosynthesis, wax, cutin, suberin metabolism, tryptophan metabolism, and phenylpropanoid biosynthesis genes. The expanded gene families classified by KEGG pathway identified genes for linoleic acid metabolism, phenylalanine metabolism, alpha-linolenic acid metabolism and diterpenoid biosynthesis, starch and sucrose metabolism, pentose and glucuronate interconversions, and phenylpropanoid biosynthesis.

Collinearity analysis and WGD

Paralogous genes were identified in chayote and pumpkin through genome collinearity analysis. This identified 41,018 collinear genes, accounting for 73.11% of the total gene number (56,010) (Fig. 5), which suggested a high degree of conserved gene order in the two species, although there were significant rearrangements. The collinearity analysis identified

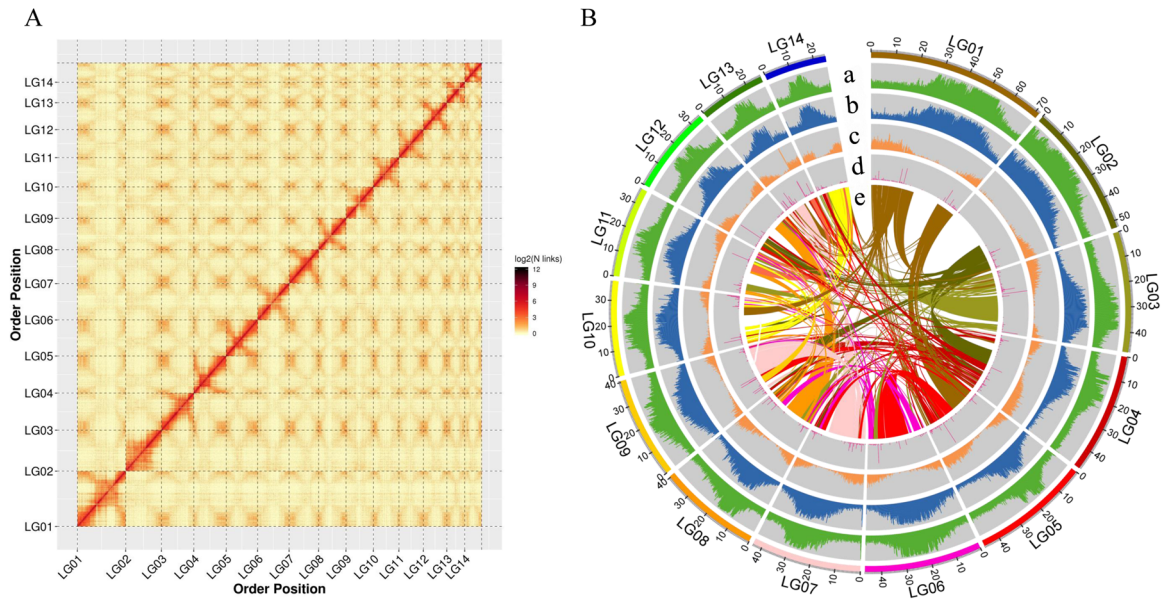

Fig. 2 Chayote genome information. A Hi–C heat map of chromosome interactions; **B** overview of chayote genomic features. **a** GC content; **b** repeat sequence coverage; **c** gene density; **d** noncoding RNA density; and **e** chromosome collinearity

Table 2 Genome annotation statistics

Parameter	Number	Percentage (%)
Total repetitive sequences	957,675	65.94
Protein-coding genes	28,237	–
miRNA	101	–
rRNA	298	–
tRNA	1873	–
Pseudogenes	1085	–
All functional annotations	27,469	97.28
GO_annotations	15,018	53.19
KEGG_annotations	9124	32.31
KOG_annotations	14,847	52.58
TrEMBL_annotations	27,435	97.16
Nr_annotations	27,451	97.22
Unannotated	768	2.72

a mass of synonymous gene blocks in chayote and pumpkin. The comparison of chayote and pumpkin indicated that there have been far more interchromosomal rearrangement events than between chayote and snake gourd (72.2%) (Fig. 5). With the exception of *peroxidase 45-like* (EVM0001323.1), the genes indicated in Fig. 5B are related to phytohormones, such as *auxin response factor 17* (EVM0017908.1), *auxin response factor 9* (EVM0019236.1), *auxin-induced protein AUX22-like* (EVM0027070.1), *1-aminocyclopropane-1-carboxylate synthase* (EVM0000580.1), *1-aminocyclopropane-1-carboxylate oxidase 5-like* (EVM0026245.1), and *ethylene-responsive transcription factor 3-like* (EVM0008801.1), which were present in all three fruit species.

WGD events are of great significance in generating species diversity during evolution. It can be deduced from the Ks and 4DTv distribution, which in chayote had peaks at approximately 0.314 and 0.102 (Figs. 5 and S7), which indicated that a paleoploidization (WGD) event occurred recently, as proposed by others[31]. These WGD events are shown by dot plots of paralogs in Fig. 5A. An ancient WGT (γ) event (approximately 130–150 Mya)[32,33] that occurred after the divergence of monocotyledons and dicotyledons[34], was confirmed for grape, towel gourd, and snake gourd by the peak shown in Fig. 5C. From the divergence date of recent WGDs, such as those in cucurbita[20] (4DTv = 0.12, 30 ± 4 Mya) and soybean[35] (4DTv = 0.057, 13 Mya), we estimated that the chayote WGD event occurred 25 ± 4 Mya ($T = D/2\mu$). According to the evolutionary tree (Fig. 4C), chayote diverged from snake gourd (27–45 Mya) and towel gourd (29–55 Mya), which suggested that the WGD of chayote occurred after its differentiation from towel gourd and snake gourd. No WGD event was observed previously for towel gourd[21,36] or snake gourd[37], consistent with our results.

Transcription and metabolism in developing chayote fruit

Transcriptomics and metabolomics were used to study changes during fruit development. Correlations between

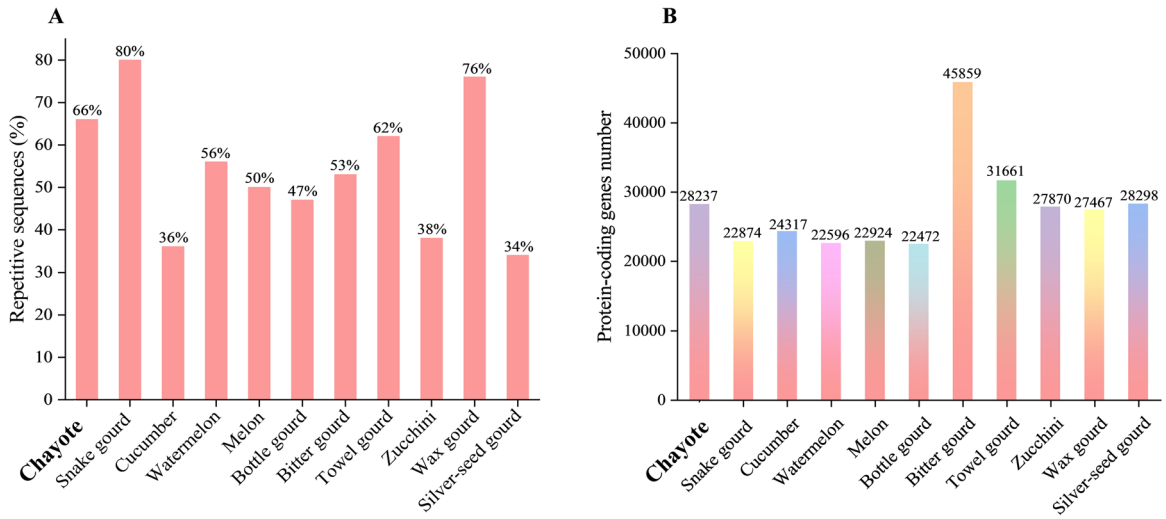

Fig. 3 Comparison of chayote with other Cucurbitaceae genome assemblies. A Percentage of repetitive sequences in chayote and ten other Cucurbitaceae; **B** comparison of numbers of protein-coding genes in chayote and ten other Cucurbitaceae (snake gourd[37], cucumber[17], watermelon[49], melon[18], bottle gourd[22], bitter gourd[50], towel gourd[21], zucchini[20], wax gourd[23], silver-seed gourd[53]). The colors in **B** indicate different species

samples in transcriptomics and orthogonal partial least squares-discriminant analysis (OPLS-DA) assessment of metabolomics were checked and compared with differential expression analysis (Fig. S8). A total of 385 differentially expressed genes (DEGs) were identified in the 3 days vs. 6 days comparison (Fig. 6), consisting of 118 upregulated genes and 267 downregulated genes; similarly, 34 different metabolites were identified as differentially abundant, including 12 that decreased and 22 that increased, and these were annotated with KEGG and GO terms (Figs. 7 and S9, Tables S6 and S7). In the comparison of 6 days vs. 9 days, a total of 57 different metabolites and 1033 DEGs were identified, consisting of 464 upregulated genes and 569 downregulated genes and 8 decreased and 49 increased metabolites (Fig. 7, Tables S6 and S7, Fig. S10). In the 3 days vs. 9 days comparison, 48 different metabolites were found, of which 40 were upregulated and 8 were downregulated (Fig. 7, Tables S6 and S7). A total of 3281 DEGs, consisting of 1606 upregulated and 1675 downregulated genes, were clustered and annotated with KEGG and GO terms (Figs. S11 and S12).

Among the three comparison groups, transcripts of genes involved in plant hormone synthesis and response were markedly changed in chayote fruit. *Auxin-induced protein 22D-like* (*AUX22D/22B/22*) and *auxin-responsive protein IAA13-like* (*IAA13/9*) were obviously upregulated in the 3 days vs. 6 days comparison. In the 6 days vs. 9 days comparison, transcripts of *auxin-responsive protein SAUR50-like* (*SAUR50*) increased 23-fold; *auxin-responsive protein IAA13-like* (*IAA13/IAA11*), *auxin-induced protein 22B-like* (*AUX22B*), and *gibberellin 20 oxidase 1-like* (*GB*) showed similar trends, while *1-aminocyclopropane-1-carboxylate synthase 7* (*ACS7*), *1-aminocyclopropane-1-carboxylate oxidase homolog 6* (*ACO6*), and *abscisic acid-insensitive 5-like protein 6* (*ABA5*) were downregulated. In the 3 days vs. 9 days comparison group, transcript levels of *ACS7*, *ACO6*, and *ABA5* also decreased. Genes encoding other enzymes associated with plant hormones were upregulated, such as *AUX22B/AUX22D*, *IAA11/IAA13/IAA14/IAA21*, and *SAUR50*. Overlaps between these comparison groups provide information about the development of chayote fruit. *AUX22B*, *AUX22D*, and *IAA13*, which are auxin growth-related factors, increased, and a role for auxin is indicated in Fig. 6E. Auxin is linked to *transport inhibitor response 1* (*TIR1*), which can direct polyubiquitylation[38]. When auxin levels increase, *AUX/IAA* proteins are targeted for proteasomal degradation, which is a key factor in the regulation of auxin signaling[39]. With the release of free active ARFs, the activation of target genes is promoted[40]. In addition, *auxin-responsive protein* (*IAA*), auxin-responsive *GH3* family genes and *SAUR* family proteins are used to regulate fruit growth and development. *IAA* and *GH3* gene families were up and downregulated, respectively, and *SAUR* family proteins expression decreased from 3 to 9 days, which was correlated with chayote fruit enlargement. Transcripts for *ACS7* and *ACO6*, which together catalyze the biosynthesis of ethylene[41], decreased from 6 to 9 days, indicating a likely reduction in ethylene production at this stage.

mRNAs for several TFs changed significantly in the comparisons between different stages of fruit development. In the 3 days vs. the 6 days comparison, *bHLH94*,

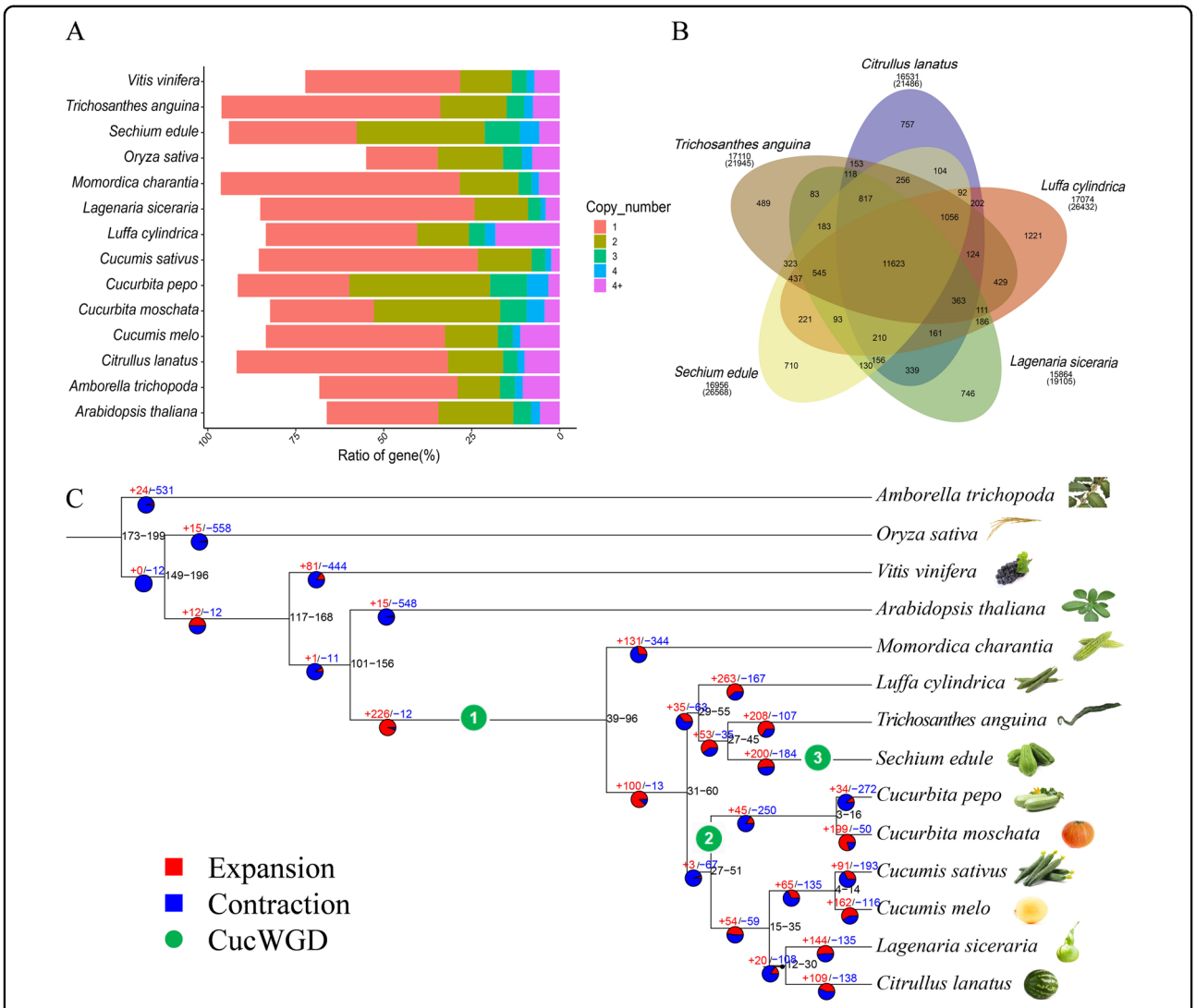

Fig. 4 Analysis of gene families and phylogeny of chayote and other related plant genomes. A Gene copy number distribution in chayote and 13 other sequenced plant genomes; **B** Venn diagram representing the clusters of gene families in chayote and four related plants (*Trichosanthes anguina, Luffa cylindrica, Citrullus lanatus, Lagenaria siceraria*); **C** phylogenetic tree and gene family expansions/contractions in chayote and 13 other species (*Cucumis sativus*, *Citrullus lanatus*, *Luffa cylindrica*, *Lagenaria siceraria*, *Cucurbita moschata*, *Trichosanthes anguina*, *Momordica charantia*, *Cucurbita pepo*, *Cucumis melo*, *Arabidopsis thaliana*, *Vitis vinifera*, *Amborella trichopoda*, and *Oryza sativa*) based on a concatenated alignment of 832 single-copy protein sequences. The tree is rooted with *A. trichopoda* as the outgroup. Red represents expanded gene families, blue represents contracted gene families, and green indicates Cucurbit-common whole-genome duplication (CucWGD)

MYB3R-1-like and *NAC TF* (*NAM-2*) increased, whereas ethylene-responsive TFs (*ERF109/017/4/11/ERF1B/AP2-1*), *MYB* (*MYB108/44/24*), *MYC* (*MYC3/2*), *WRKY* (*WRKY40/46/2*), *bHLH 93*, *auxin response factor 9* (*ARF9*), *TF HBP-1b*, *GATA TF 9-like* were downregulated. The transcription of *MYC2*, which is involved in jasmonic acid signaling, decreased. In the 6 d vs 9 d comparison, transcripts for TFs, such as *bHLH* (*bHLH118/91/82/69*), *ERF* (*ERF105/98/106/RAP2-3*), *bZIP* (*bZIP11/44*), probable *WRKY* (*WRKY49/57*), and *MYB* (*MYB59/48*) were upregulated, and those for other *ERFs* (*ERF60/110/11/61/AIL1/5/6*), *MYB* (*MYB41/44/111*), *TCP* (*TCP2/17*), *bHLH149*, *PIF3*, *GTE7*, and *VOZ1* were decreased. In the 3 days vs. 9 days comparison, downregulated TFs included *WRKY* (*WRKY 22/12/7/1/34*), *PIF3/PIF5*, *MYC2*, *TGA1/TGA9*, *ERF* (*ERF11/3/4/53/60/61/106/110/113/RAV1/2/AIL1/5/6/ANT/RAP2-1/2-4/2-7*), *MYB* family TF (*PHL11*), *bHLH* (*bHLH13/112/149/30/69/74/78/93/102/108/111/1R1/20/24/30/41/44*), and *NAC1/NAC25*. The upregulated TFs included *bZIP11/bZIP14*, *ERF* (*ERF98/105/106/RAP2-10/2-11/2-1/2-3*), *MYB* family TFs (*APL/PHL4*), *NAC* (*NAM-2*), *WRKY* (*WRKY 21/23/4/57*), and *bHLH* (*bHLH110/118/143/155/51/68/78/79/82/91/3R-1/48/59*).

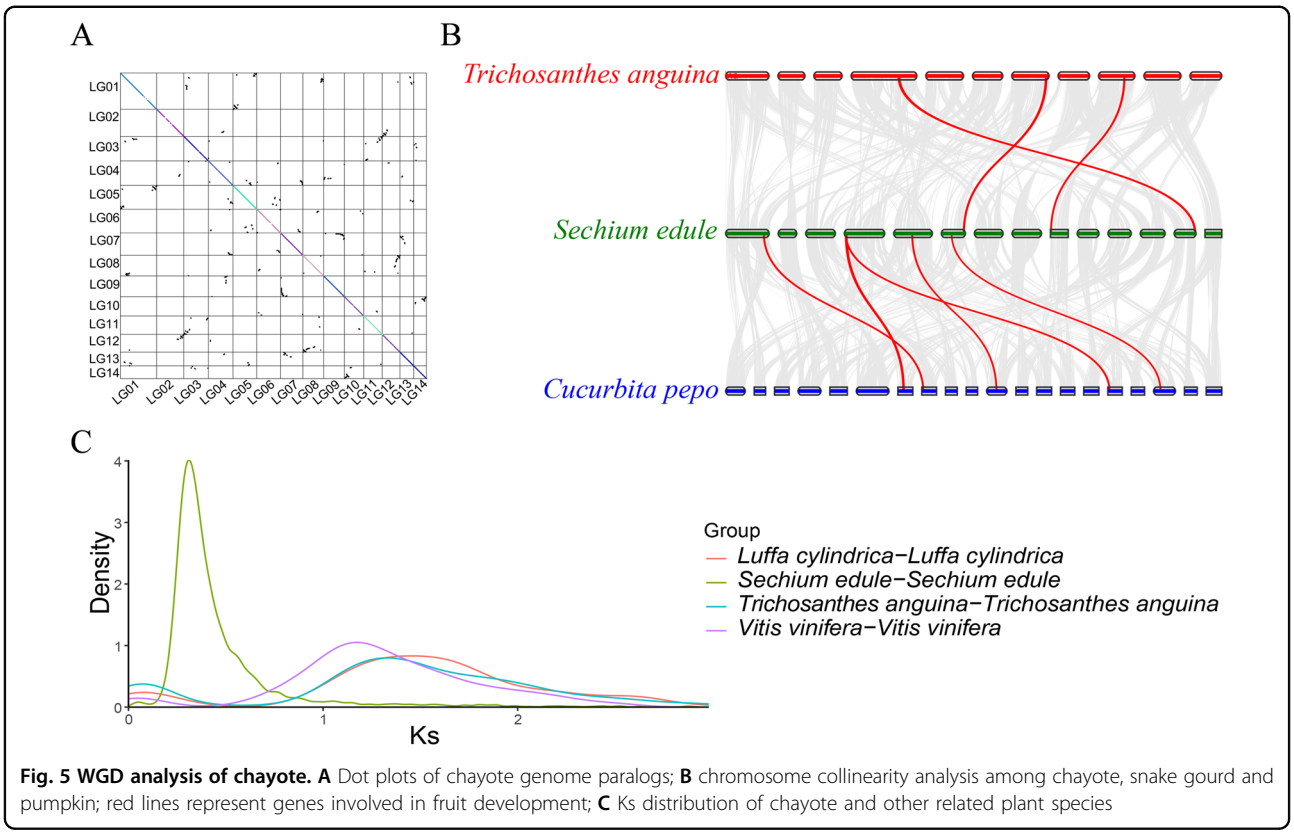

Fig. 5 WGD analysis of chayote. A Dot plots of chayote genome paralogs; **B** chromosome collinearity analysis among chayote, snake gourd and pumpkin; red lines represent genes involved in fruit development; **C** Ks distribution of chayote and other related plant species

bHLH and *MYB* are significant TF families for controlling the biosynthesis of isoflavonoids and flavonoids[42]. Phenylpropanoid biosynthesis pathway was the upstream part of flavonoid biosynthesis, and the transcript levels of several of these genes changed markedly in the 6 days vs. 9 days comparison group (Fig. 6H); some of these genes were associated with the production of lignin and phenolic compounds rather than flavonoids. Through *phenylalanine ammonia-lyase* (*PAL*), cinnamic acid is converted to cinnamoyl-CoA and p-coumaroyl-CoA by *4-coumarate-CoA ligase* (*4CL*) and *trans-cinnamate 4-monooxygenase* (*CYP73A*), which has been confirmed to participate in the biosynthetic pathway leading to flavonoids[43]. Genes encoding enzymes involved in the production of alcohols, such as *cinnamoyl-CoA reductase* (*CCR*), *caffeic acid 3-O-methyltransferase* (*COMT*), and *ferulate-5-hydroxylase* (*F5H*, *CYP84A*), were upregulated. Peroxidase (POD) is involved in lignin production. The concentration of many metabolites increased, such as hispidulin, luteolin-4′-O-glucoside, kaempferol-7-O-glucoside, quercetin-3-O-glucoside, luteolin-C-rhamnosyl-glucoside, luteolin-7-O-rutinoside, kaempferol-3-O-robinobioside (biorobin), and luteolin-3'-O-glucoside in the 3 d vs 6 d comparison. Furthermore, in the 6 days vs. 9 days comparison, luteolin-7-O-glucoside (cynaroside or luteoloside) (23-fold increase), luteolin-4′-O-glucoside (11-fold increase), luteolin-3′-O-glucoside (10-fold increase), isosinensetin (10-fold increase), kaempferol-7-O-glucoside (9.5-fold increase), kaempferol-3-O-robinobioside (biorobin) (8.9-fold increase), pratensein, nobiletin, hispidulin, quercetin-3-O-rhamnoside, luteolin-7-O-rutinoside, kaempferol-3-O-neohesperidoside, luteolin-6-C-glucoside (isoorientin) and diosmetin-7-O-galactoside concentrations increased. Notable increases in metabolites from 3 to 9 days samples included luteolin-7-O-glucoside (cynaroside) (151-fold), kaempferol-3-O-neohesperidoside (39-fold), kaempferol-3-O-robinobioside (biorobin) (38-fold), luteolin-7-O-rutinoside (36-fold), luteolin-C-rhamnosyl-glucoside (34-fold), kaempferol-7-O-glucoside (33-fold), luteolin-4′-O-glucoside, luteolin-3′-O-glucoside (33-fold), quercetin-3-O-glucoside (32-fold), hispidulin (12-fold), pratensein (11-fold), isosinensetin (9-fold), diosmetin-7-O-galactoside (8-fold), isoorientin (5.6-fold), and nobiletin (2.5-fold). Kaempferol-3-O-neohesperidoside, kaempferol-3-O-robinobioside (biorobin), kaempferol-7-O-glucoside, luteolin-7-O-rutinoside, luteolin-4′-O-glucoside, luteolin-3′-O-glucoside and hispidulin were common across the three comparison groups. Luteolin-C-rhamnosyl-glucoside and quercetin-3-O-glucoside remained at similar concentrations from 6 to 9 days. In contrast, luteolin-7-O-glucoside (cynaroside), isosinensetin, pratensein, nobiletin, luteolin-6-C-glucoside (isoorientin), and diosmetin-7-O-galactoside increased from 6 to 9 days.

Some identified gene transcripts were likely associated with fruit quality. In the comparison of 3 days vs. 6 days,

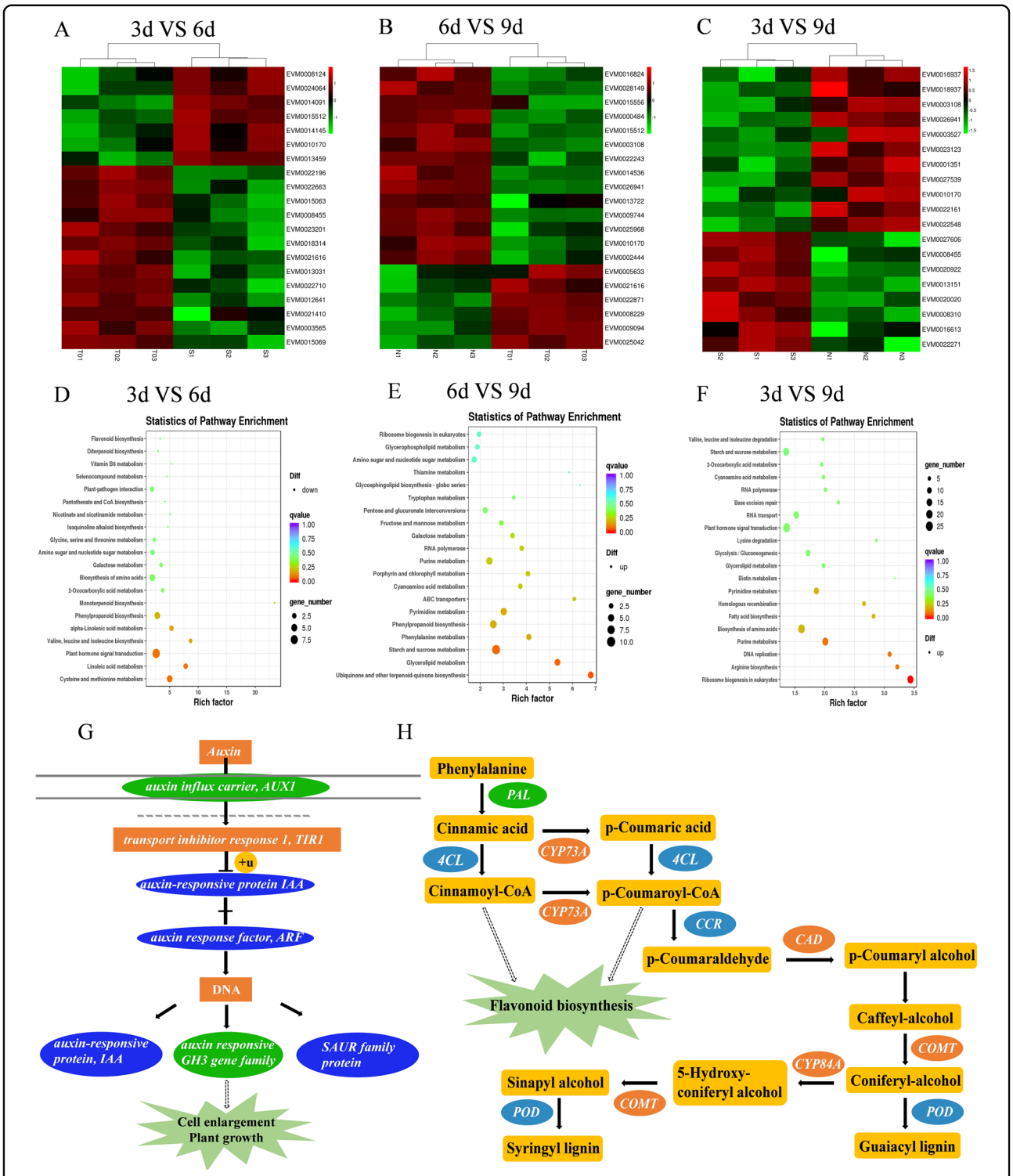

Fig. 6 Transcriptome KEGG annotation and enrichment map for genes expressed during chayote fruit development. A Heat map of major genes differentially expressed in fruit (3 days vs. 6 days); **B** heat map of major genes differentially expressed in fruit (6 days vs. 9 days); **C** heat map of major genes differentially expressed in fruit (3 days vs. 9 days); **D** KEGG annotation of transcripts decreased from 3 days to 6 days fruit; **E** KEGG annotation of transcripts increased from 6 days to 9 days fruit; **F** KEGG annotation of transcripts increased from 3 days to 9 days fruit; **G** changes in gene expression related to plant hormone signal transduction pathways in 3 days vs. 9 days; green represents decreased, blue represents decreased and increased; **H** changes in expression of gene of phenylpropanoid biosynthesis in 6 days vs. 9 days, orange represents increased, green represents decreased, blue represents decreased and increased, yellow represents unchanged

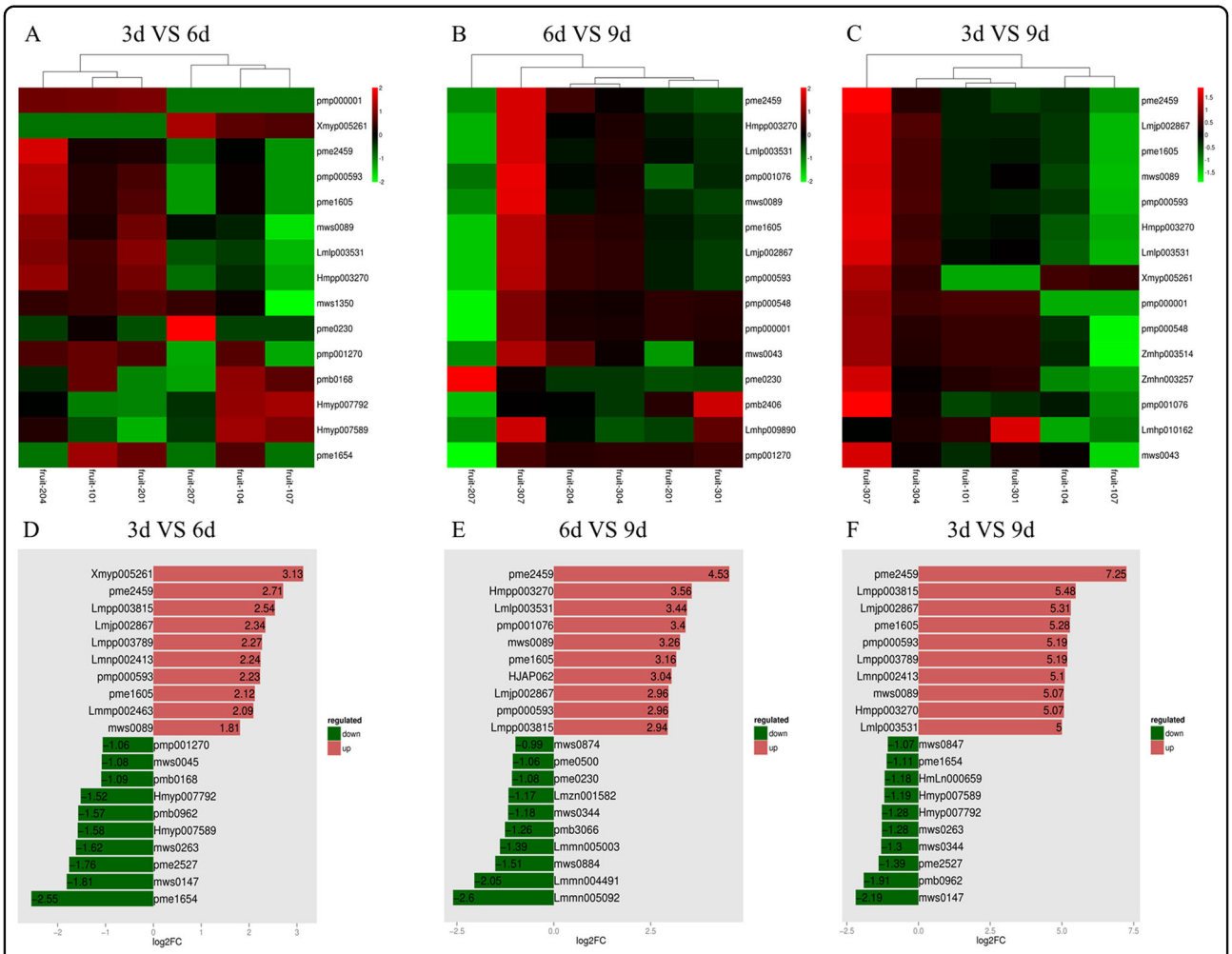

Fig. 7 Changes in metabolism during fruit development. A Heat map showing major changes in metabolites in 3 days vs. 6 days fruit; **B** heat map showing major changes in metabolites in 6 days vs. 9 days fruit; **C** heat map showing major changes in metabolites in 3 days vs. 9 days fruit; **D** top 20 greatest fold-changes in metabolites in the comparison of in 3 days vs. 6 days; **E** top 20 greatest fold-changes in metabolites in the comparison of 6 days vs. 9 days; **F** top 20 greatest fold-changes in metabolites in the comparison of 3 days vs. 9 days

peroxidase 72-like (*POD*), *polyphenol oxidase* (*PPO*), *chlorophyllase-1* (*CHL1*), *linoleate 13S-lipoxygenase 2-1* (*LOX2-1*), *beta-amyrin 11-oxidase-like* and *chalcone synthase 2* (*CHS2*) were downregulated. In contrast, *histidine kinase 4-like* (*HK4*), *monogalactosyl diacylglycerol synthase 2* (*MGDG2*), *7-hydroxymethyl chlorophyll a reductase* (*HCAR*), *cytochrome P45090B1*, and *two-component response regulator ARR11* (*ARR11*) were upregulated. For metabolomics analysis of 3 days vs. 6 days, 2-hydroxyhexadecanoic acid, syringic aldehyde, and isorhamnetin-3-O-(6′-p-coumaroylglucoside) were increased. In the comparison of 6 days vs. 9 days, *9-cis-epoxycarotenoid dioxygenase* (*NCED2*), *peroxidase 27-like* (*POD27/POD66*), *LOX2-1*, *beta-carotene hydroxylase 2* (*CHY2*), *gibberellin 2-beta-dioxygenase 8-like protein* (*GB8*), *9-cis-epoxycarotenoid dioxygenase* (*NCED3*), *phenylalanine ammonia-lyase 5* (*PAL5*), *phenylalanine ammonia-lyase-like* (*PAL*), *zeaxanthin epoxidase* (*ZEP*), *probable carotenoid cleavage dioxygenase 4* (*CCD4*), *peroxidase 2-like* (*POD2*), and *carotenoid 9,10-cleavage dioxygenase 1* (*CCD1*) were increased, together with transcripts for the cell wall modifying enzymes *polygalacturonase* (*PG*) and *pectinesterase 2* (*PE*). *Peroxidase* (*POD*) is involved in cell wall lignin formation (Fig. 6F) and may be associated with chayote fruit texture. Other metabolites that also increased included lysoPC (20:3, 17:0, 17:1, 17:2, and 16:1), lysoPC (18:1, 17:0, 20:2, 20:3, 16:1, and 14:0) (2*n* isomer)*, lysoPE 15:1 and lysoPE (15:1, 16:1, 17:1, 18:1, 20:3, and 20:2) (2*n* isomer)*, which are associated with cell membrane structure[44] and may be related to the rapid enlargement of chayote fruit. Comparison of the 3 days vs. 9 days transcriptome changes identified transcripts involved in cell wall structure and fruit texture, such as *expansin-like B1*, *glucan endo-1,3-*

beta-glucosidase 12, *phenylalanine ammonia-lyase 5* (*PAL5*), and *glutamate dehydrogenase 2*, which were downregulated. Other gene transcripts, such as *POD2*, *ARR12*, *GB8*, *NCED2* and *NCED3*, also decreased. Several transcripts were increased, including *expansin-like B1*, which could be involved in inducing plant cell wall extension[45,46], consistent with the rapid enlargement of chayote fruit. In addition, *HK4* and *CCD4*, which participate in ethylene signal transduction[47] and flavonoid production[48], respectively, also increased. Unsaturated acids were changed between the 3 days vs. 9 days metabolomes, besides flavonoids, isoflavones, lysoPC, and lysoPE.

Association analysis between the metabolome and transcriptome can aid in the understanding of transcriptional regulation mechanisms controlling metabolic pathways. According to this analysis, plant hormone signal transduction, alpha-linolenic acid metabolism, 2-oxocarboxylic acid metabolism, arginine biosynthesis, and glutathione metabolism were enriched in the 3 days vs. 6 days comparison, whereas flavonoid biosynthesis and accumulation were increased in the 6 days vs. 9 days comparison.

Discussion

Chayote is popular for its appearance, nutrition and palatability and is a significant economic crop. This study reports the first high-quality genome assembly of 14 chayote chromosomes. The genome size is 606.42 Mb, which is similar to that of *L. cylindrica*[21] (669 Mb) but smaller than those of *T. anguina*[37] (919.8 Mb) and *B. hispida* (859 Mb)[23]. Other members of the Cucurbit have much smaller genomes, such as *C. sativus*[17], *C. melo*[18], *C. lanatus*[49], *C. pepo*[20], and *M. charantia*[50]. Approximately 401.08 Mb (65.94%) of repetitive sequences were predicted, which was less than the amount in the *B. hispida* genome and greater than that in the *C. melo* genome[18,23]. A total of 27,469 protein-coding genes were annotated, which is similar to the numbers in wax gourd and pumpkin[23,51]. This genome assembly will underpin further deep molecular-level research, trait selection, and evolutionary studies in Cucurbitaceae.

This study provides insights into WGD events and chayote evolution. WGD generates gene homologs and thus is of great significance during the diversification of species and the acquisition of new functions. As others predicted, Cucurbitaceae underwent four WGD events. First, cucurbit-common tetraploidization (CucWGD1 in Fig. 4C) occurred shortly after core-eudicot-common hexaploidy (ECH, 115–130 Mya)[52]. Second, pumpkin[51] (*Cucurbita maxima* and *C. moschata*), zucchini[20] (*C. pepo*) and silver-seed gourd[53] (*Cucurbita argyrosperma*) were confirmed to have undergone a WGD event (CucWGD2 in Fig. 4C). Third, our data indicated that chayote underwent an additional WGD event at approximately 25 ± 4 Mya, denoted as CucWGD3 (Fig. 4C), which requires further exploration. Interestingly, the genome collinearity percentage between chayote and pumpkin was slightly higher than that between chayote and snake gourd (Fig. 5B). This may indicate that chayote generated more new genes, perhaps to adjust to a changed environment, after WGD occurred. A high degree of genome collinearity suggested that chayote contains conserved ancestral Cucurbitaceae genes, as is the case in pumpkin. Considerable changes have occurred during the evolution of cucurbitaceous species, as seen from the different shapes and lengths of chayote, snake gourd and towel gourd, which are closely related, according to the evolutionary tree (Fig. 4C).

Transcripts of genes affecting various metabolic pathways were found to increase during chayote fruit development in our study. A number of these factors affect pigment content, such as *HCAR* (7-hydroxymethyl chlorophyll a reductase), which, together with *CHL1*, is required for chlorophyll a synthesis[54] and thus may affect chayote chlorophyll content. Several of the transcripts encoded enzymes that are involved in carotenoid metabolism. Beta-carotene hydroxylase (*CHY2*) can influence carotenoid content. *CCD1* and *CCD4* encode enzymes that cleave various carotenoids to yield β-ionone and α-ionone in other plants[48,55,56] and are likely to affect fruit carotenoid content and flavor. It is probable that *NCED*2 and *ZEP* also participate in regulating chayote fruit pigments[57] and abscisic acid biosynthesis[58]. Genes related to other hormones also showed increased expression, including *AUX*, *IAA* and *HK4*, which may be related to the major growth of chayote fruit[59]. Cellulose synthase-like protein B4 is associated with cell wall complexes[60] and, together with polygalacturonase, pectinesterase may influence chayote cell wall properties and texture during fruit development. Furthermore, *13-LOX2-1* may be involved in chayote fruit defense[61], although some LOX genes are also involved in flavor generation.

Several isoflavonoid and flavonoid compounds increased in amount, such as nobiletin, hispidulin, isosinensetin, quercetin-3-O-rhamnoside, luteolin-4′-O-glucoside, luteolin-3′-O-glucoside and luteolin-7-O-rutinoside. Quercetin[62], nobiletin and hispidulin are antioxidants with superoxide anion-scavenging activity, and their increased content in chayote may contribute important dietary health benefits[63,64]. A schematic showing major gene expression changes during chayote fruit development was constructed based on information from the transcriptome and metabolome (Fig. 8). Further in-depth investigation of the action of the genes identified in this study may explain specific aspects of chayote fruit texture, pigment, growth, flavor, and antioxidant properties.

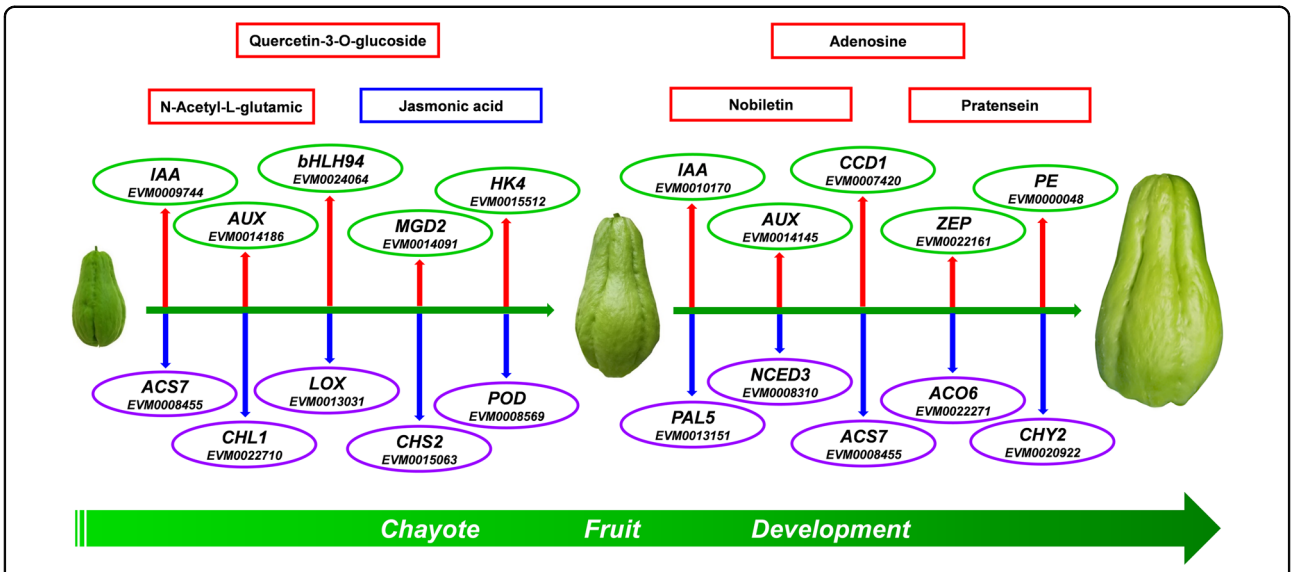

Fig. 8 Schematic showing changes in transcripts during chayote fruit development. From left to right, fruit at 3, 6, and 9 days. The upward-facing arrows (red) indicate transcripts that increase with time. The downward-facing arrows (blue) represent transcripts that decrease with time. Metabolic components are circled in red (upregulated) or blue (downregulated)

Materials and methods

Chayote sample collection and genomic sequencing

Fresh chayote leaves were collected from Jianshui in Yunnan Province for genome sequencing. DNA was extracted from young chayote leaves, and two 350 bp libraries were constructed. The library was sequenced (150 bp reads at each end, PE150) using an Illumina sequencer. The raw information was assessed for GC distribution statistics and Q20 and Q30 quality value evaluations and then filtered to obtain clean reads that were used for genome size evaluation, genome assembly, GC content statistics, heterozygosity statistics, and post-assembly evaluation.

A Nanopore library was constructed and used for Nanopore third-generation sequencing in five steps: (1) large DNA fragment isolation, (2) fragment repair, (3) connecting reactions, (4) quantitative detection, and (5) library construction. Finally, single-molecule real-time sequencing was carried out on the PromethION sequencer to obtain the raw data prior to error correction to obtain high accuracy data. De novo genome assembly was performed by combining three strategies: initially, in the correction step, longer reads were selected by Canu[65] (available at https://github.com/marbl/canu, v1.5) with the settings 'genomeSize=1000000000' and 'corOutCoverage=50', subsequently, overlapping raw reads were identified through the high-sensitivity overlapper MHAP (mhap-2.1.2, option 'corMhapSensitivity=low/normal/high'), subsequently, error correction was performed by the falcon_sense method (option 'correctedErrorRate=0.025'); followed by Smartdenovo, error correction using racon[66] software (https://github.com/isovic/racon) and adjustment by Pilon[67] software (v1.22, available at https://github.com/broadinstitute/pilon, with the parameters ('--mindepth 10 --changes --threads 4 --fix bases'). The assembled results were assessed by evaluation of the ratio of the Illumina sequencing reads and the evaluation of BUSCO integrity. BUSCO[30] v 2.0 (eudicotyledons_odb9 database) was used to validate the genome completeness and gene set completeness of the draft genome sequences.

Hi–C fragment libraries were constructed with 300–700 bp insert sizes and sequenced by the Illumina platform after five steps: (1) cell crosslinking, (2) endonuclease digestion, (3) end repair, (4) DNA cyclization, and (5) DNA purification and capture. Qubit 2.0 and Agilent 2100 instruments were used to detect library concentration and insert size. First, raw reads were trimmed, and low-quality PE reads were removed to obtain clean data. The clean Hi–C reads were aligned to the assembly results with bwa aligner[68] (version: 0.7.10-r789) after a trim of clean reads at the putative Hi–C junctions. Only alignable read pairs with mapping quality greater than 20 were reserved for further analysis. Invalid read pairs, including dangling-end, self-circle, re-ligation, and "dumped" products, were filtered by HiC-Prov2.8.1. LACHESIS[69] software was used for grouping, sequencing and orientation of genome sequences, and the assembly results were evaluated. Parameters for running LACHESIS were as follows: CLUSTER_MIN_RE_SITES, 64; CLUSTER_MAX_LINK_ DENSITY, 2; CLUSTER_NONINFORMATIVE_RATIO, 2; ORDER_MIN_N_RES_IN_TRUN, 15; and ORDER _MIN_N_RES_IN_SHREDS, 15.

Gene prediction and function annotation

Based on the principles of structure prediction and de novo prediction, we constructed the repeat sequence database of the chayote genome through LTR_FINDER[70] and RepeatScout[71]. A database of repeat elements in the chayote sequence was generated by PASTEClassifier[72] and then merged with the Repbase database[73] to generate the final database of repetitive sequences. The repetitive sequences of the chayote genome were predicted by RepeatMasker[74] software based on the constructed repetitive sequence database.

Coding gene prediction analysis in the chayote scaffold sequences was carried out using three different methods: de novo prediction, homology-based species prediction and UniGene prediction. Then, EVM[73] v1.1.1 software was used to integrate the prediction results. Genscan[75], Augustus[76] v2.4, GlimmerHMM[77] v3.0.4, GeneID[78] v1.4, and SNAP[79] were used for *de nove* prediction; GeMoMa[80] v1.3.1 was used for homologous species prediction; Hisat[81] v2.0.4 and Stringtie[82] v1.2.3 were used for assembly based on reference transcripts; TransDecoder v2.0 and GeneMarkS-T[83] v5.1 were used for gene prediction; PASA[84] v2.0.2 was used for the prediction of UniGene sequences without reference assembly based on transcriptome data; EVM v1.1.1 was used to integrate the prediction results obtained by the above three methods; and PASA v2.0.2 was used for modification. In addition, we also predicted different noncoding RNAs. Whole genome comparison and recognition of microRNAs and rRNAs were carried out with Blastn based on the Rfam[85] database; tRNAs were identified by tRNAscan-SE[86]. Pseudogene prediction was also performed. Using the predicted protein sequence, through BLAT[87] comparison, homologous gene sequences (possible genes) were identified in the genome, and we then used genewise[88] to find premature termination codons and frameshift mutations in the gene sequences to identify the pseudogenes.

The predicted gene sequences were compared with NR[89], KOG[90], GO, KEGG[91], TrEMBL[92], and other functional databases. Databases were used to compare gene protein sequences by BLAST[93] v2.2.31 (E value $\leq 1 \times 10^{-5}$), and gene functions were annotated by KEGG pathway annotation analysis. KOG functional annotation analysis, GO functional annotation analysis, and other gene functional annotation analyses were carried out to obtain final annotations.

Comparative analysis of genomes between species

Using OrthoFinder v2.3.7 software[94], the protein-coding sequences in the chayote genome and genomes from 13 other species were compared. The PANTHER V15 database[95] was used to annotate the obtained gene families. Finally, GO and KEGG enrichment analyses were carried out by using clusterProfiler v3.14.0[96] to identify the gene families unique to chayote.

The evolutionary tree was constructed using the maximum likelihood method by IQ-TREE v1.6.11[97] software and single-copy protein sequences, with the number of bootstrap replicates set to 1000. For the evolutionary tree, we set the outer group as *A. trichopoda* to obtain a rooted tree. MCMCTREE, a software package in PAML v4.9i[98], was used to calculate the divergence time. The number of Markov chain iterations was set as follows: burnin 500000, sampfreq 10, nSample 5000000.

According to the evolutionary tree results with differentiation time and gene family clustering by CAFE v4.2 (Computational Analysis of gene Family Evolution) software[99], the number of gene family members of each branch's ancestors was estimated by a birth mortality model to predict the contraction and expansion of gene families of species relative to their ancestors ($p < 0.05$). We identified the expanded and contracted gene families in chayote and annotated them with PANTHER. GO and KEGG enrichment analyses were carried out with clusterProfiler.

WGD event and collinearity analysis

Diamond v0.9.29.130[100] was used to compare the gene sequences between two species and identify similar gene pairs (E value $< 1 \times 10^{-5}$, C score > 0.5, JCVI software[101] was used to filter the C score). Then, according to the gff3 document, MCScanX[102] was used to determine whether similar gene pairs are adjacent on chromosomes and finally, collinear gene blocks were obtained.

WGDs are events in which the genome is doubled. At present, the Ks (synonymous mutation rate) method and 4DTv (fourfold synonymous third-codon transversion rate) methods are commonly used to identify WGDs; here, wgd v1.1.0 software[103] and a custom script (https://github.com/JinfengChen/Scripts) were used to identify WGD events in chayote.

Transcription analysis

Fresh chayote fruit samples at 3, 6, and 9 days were collected for transcription analysis, and three biological replicates were performed for each stage. The cDNA libraries were obtained by polymerase chain reaction (PCR) enrichment. After the library was checked for quality by Q-PCR, the Illumina platform was used for sequencing. Low-quality and adapter reads were removed to obtain clean data, which were used for sequence alignment with the specified reference genome. The transcriptome was assembled using StringTie[82]. Differential expression analysis was performed between the different sample groups. Pearson's correlation coefficient (r) was used for repeatability assessment[104]. DESeq2[105]

was used for differential expression analysis between sample groups to obtain the DEG sets between two biological samples. Hierarchical clustering analysis was carried out for the screened DEGs to find groups of genes with the same or similar expression patterns. For the detection of DEGs, fold change > 2 and false discovery rate (FDR) < 0.01 were used as cutoff values. As a screening standard, fold change (0.01) represents the FDR between two samples with a corrected p value for significant differences. For functional annotation and enrichment analysis of DEGs, gene function annotation was performed as described for gene prediction and functional annotation above.

Metabolites analysis

Freshly collected chayote fruit samples were used for metabolome analysis. The freeze-dried samples were crushed into powder prior to ultra performance liquid chromatography–mass spectrometry (MS)/MS[106,107] analysis. Analyst 1.6.3 software was used to process mass spectrum data. Based on the local metabolic database, the metabolites contained in the samples were quantitatively and qualitatively analyzed by MS[108,109]. To determine the repeatability of metabolite extraction and detection, one QC sample was inserted per ten samples. The total ion current diagram was overlapped and analyzed. Principal component analysis[110], cluster analysis[96] and Spearman rank correlation (R) were used for metabolome data assessment. OPLS-DA[111] analysis was performed for each difference group. Subsequently, the differential multiple, p value of the t test and VIP value of the OPLS-DA model were combined to screen differentially accumulated metabolites[111]. The screening criteria were FC > 2, p value < 1 and VIP > 1.

Transcriptome and metabolism conjoint analysis

The results of metabolome analysis were combined with the results of transcriptome analysis, and the DEGs and differentially accumulated metabolites in the same group were simultaneously mapped to the KEGG pathway map[112]. Genes and related metabolic pathways were identified for analysis and data screening using a p value < 0.05 to identify significant relationships.

Acknowledgements

This work was supported by grants from the Beijing Municipal Science and Technology Commission (Z191100008619004 and Z191100004019010), the Key Project "Science and Technology Boost the Economy 2020", the Special Innovation Ability Construction Fund of Beijing Academy of Agricultural and Forestry Sciences (20180404 and 20200427), the China Agriculture Research System Project (CARS-23), the National Key Research and Ripening Program of China, Collaborative Innovation Center of Beijing Academy of Agricultural and Forestry Sciences (201915), the Young Investigator Fund of Beijing Academy of Agricultural and Forestry Sciences (202016), the National Natural Science Foundation of China (31772022), and the Natural Science Foundation of Beijing (6182016).

Author details

[1]Key Laboratory of Vegetable Postharvest Processing, Ministry of Agriculture, Beijing Key Laboratory of Fruits and Vegetable Storage and Processing, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China) of Ministry of Agriculture, Key Laboratory of Urban Agriculture (North) of Ministry of Agriculture, The Collaborative Innovation Center of Cucurbits Crops, Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China. [2]College of Food Science and Technology, Hebei Agricultural University, Baoding 071001, China. [3]Beijing Advanced Innovation Center for Food Nutrition and Human Health, Beijing Technology and Business University, Beijing 100048, China. [4]Department of Horticulture, Zhejiang University, Hangzhou 310058, China. [5]Beijing Academy of Forestry and Pomology Sciences, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100093, China. [6]Biomarker Technologies Corporation, Beijing 101300, China. [7]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK

Data availability

The chayote raw genome and transcriptome sequencing data are available from the NCBI under project ID PRJNA640239.

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41438-021-00487-1.

Received: 23 July 2020 Revised: 7 January 2021 Accepted: 7 January 2021
Published online: 31 January 2021

References

1. Gregorio-Cipriano, R., Gonzalez, D., Felix-gastelum, R. & Chacon, S. Neoerysiphe sechii (Ascomycota: Erysiphales): a new species of powdery mildew found on *Sechium edule* and *Sechium mexicanum* (Cucurbitaceae) in Mexico. *Botany* **98**, 185–195 (2020).
2. Vieira, E. F., Pinho, O., Ferreira, I. M. P. L. V. O. & Delerue-Matos, C. Chayote (*Sechium edule*): a review of nutritional composition, bioactivities and potential applications. *Food Chem.* **275**, 557–568 (2019).
3. Lim, T *Sechium edule. Edible Medicinal And Non-Medicinal Plants*. (Dordrecht: Springer, 2012.
4. Shiga, T. M., Goncalves Peroni-Okita, F. H., Carpita, N. C., Lajolo, F. M. & Cordenunsi, B. R. Polysaccharide composition of raw and cooked chayote (*Sechium edule* Sw.) fruits and tuberous roots. *Carbohydr. Polym.* **130**, 155–165 (2015).
5. Diaz-de-Cerio, E., Verardo, V., Fernandez-Gutierrez, A. & Maria Gomez-Caravaca, A. New insight into phenolic composition of chayote (*Sechium edule* (Jacq.) Sw.). *Food Chem.* **295**, 514–519 (2019).
6. Siciliano, T., De Tommasi, N., Morelli, I. & Braca, A. Study of flavonoids of *Sechium edule* (Jacq) swartz (Cucurbitaceae) different edible organs by liquid chromatography photodiode array mass spectrometry. *J. Agric. Food Chem.* **52**, 6510–6515 (2004).
7. Abdelnour, A. & Rocha, O. J. Genetic characterization of a collection of chayote, *Sechium edule* (Jacq.) Swartz, in Costa Rica by using isozyme markers. *Genet. Resour. Crop Evol.* **55**, 163–170 (2008).
8. Cadena-Iñiguez, J. et al. Production, genetics, postharvest management and pharmacological characteristics of *Sechium edule* (Jacq.) Sw. *Fresh Prod.* **1**, 41–53 (2007).
9. Castrom, A. C., Esguerra, E. B., Agravante, J. U. & Fernando, L. M. Physiological responses to low temperature conditioning and chitosan coating of red-fleshed dragon fruit *Hylocereus polyrhizus* (Weber) Britton & Rose. *Philipp. Agric. Sci.* **101**, 344–353 (2018).
10. Loizzo, M. R. et al. Antioxidant and carbohydrate-hydrolysing enzymes potential of *Sechium edule* (Jacq.) swartz (Cucurbitaceae) peel, leaves and pulp fresh and processed. *Plant Foods Hum. Nutr.* **71**, 381–387 (2016).
11. Rietjens, I. M. et al. The pro-oxidant chemistry of the natural antioxidants vitamin C, vitamin E, carotenoids and flavonoids. *Environ. Toxicol. Pharmacol.* **11**, 321–333 (2002).

12. Mishra, L. K. & Das, P. Nutritional evaluation of squash (*Sechium edule*) germplasms collected from Garo Hills of Meghalaya-North East India. *Int. J. Agric.* **8**, 971 (2015).
13. Chao, P. Y. et al. Antioxidant activity in extracts of 27 indigenous Taiwanese vegetables. *Nutrients* **6**, 2115–2130 (2014).
14. Rosado-Perez, J., Aguiniga-Sanchez, I., Santiago-Osorio, E. & Manuel Mendoza-Nunez, V. Effect of *Sechium edule* var. nigrum spinosum (Chayote) on oxidative stress and pro-inflammatory markers in older adults with metabolic syndrome: an exploratory study. *Antioxidants* **8**, 146 (2019).
15. Lalthansanga, J. & Samanta, A. K. Effect of feeding chayote (*Sechium edule*) meal on growth performance and nutrient utilization in indigenous pig (Zovawk) of Mizoram. *Vet. World* **8**, 918–923 (2015).
16. Ramirez-Reyes, T. I. et al. Identification of antibacterial phenolics in selected plant species from mexican cloud forest by mass spectrometry dereplication. *Chem. Biodivers.* **16**, e1800603 (2019).
17. Li, Q. et al. A chromosome-scale genome assembly of cucumber (*Cucumis sativus* L.). *GigaScience* **8**, giz072 (2019).
18. Zhang, H. et al. A high-quality melon genome assembly provides insights into genetic basis of fruit trait improvement. *Iscience* **22**, 16–27 (2019).
19. Guo, S. et al. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* **45**, 51–58 (2013).
20. Montero-Pau, J. et al. De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the Cucurbita genus. *Plant Biotechnol. J.* **16**, 1161–1171 (2018).
21. Zhang, T. et al. Long-read sequencing and de novo assembly of the sponge gourd cylindrica (L.) Roem. *Genome Mol. Ecol. Resour.* **20**, 511–519 (2020).
22. Wu, S. et al. The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. *Plant J.* **92**, 963–975 (2017).
23. Xie, D. et al. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. *Nat. Commun.* **10**, 5158 (2019).
24. Woycicki, R. et al. The genome sequence of the north-european cucumber (*Cucumis sativus* L.) unravels evolutionary adaptation mechanisms in plants. *PLoS ONE* **6**, e22728 (2011).
25. Galindo-Tovar, M. E. et al. Comparison of four DNA extraction methods on various tissues and types of *Sechium edule* (Jacq.) Sw. *Phyton Int. J. Exp. Bot.* **83**, 307–310 (2014).
26. Verma, V. K., Pandey, A., Jha, A. K. & Ngachan, S. V. Genetic characterization of chayote *Sechium edule* (Jacq.) Swartz. landraces of North Eastern Hills of India and conservation measure. *Physiol. Mol. Biol. Plants* **23**, 911–924 (2017).
27. Chen, F. et al. Genome sequences of horticultural plants: past, present, and future. *Hortic. Res.* **6**, 112 (2019).
28. Chen, F. et al. The sequenced angiosperm genomes and genome databases. *Front. Plant Sci.* **9**, 418 (2018).
29. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
30. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
31. Guo, J. et al. Phylotranscriptomics in cucurbitaceae reveal multiple whole genome duplications and key morphological and molecular innovations. *Mol. Plant* **13**, 1117–1133 (2020).
32. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
33. Salse, J. Ancestors of modern plant crops. *Curr. Opin. Plant Biol.* **30**, 134–142 (2016).
34. Jaillon, O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
35. Schmutz, J. et al. Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
36. Wu, H. et al. A high-quality sponge gourd (*Luffa cylindrica*) genome. *Hortic. Res.* **7**, 128 (2020).
37. Ma, L. et al. The genome and transcriptome analysis of snake gourd provide insights into its evolution and fruit development and ripening. *Hortic. Res.* **7**, 199 (2020).
38. Depuydt, S. & Hardtke, C. S. Hormone signalling crosstalk in plant growth regulation. *Curr. Biol.* **21**, 365–373 (2011).
39. Kelley, D. R. & Estelle, M. Ubiquitin-mediated control of plant hormone signaling. *Plant Physiol.* **160**, 47–55 (2012).
40. Guilfoyle, T. J. & Hagen, G. Auxin response factors. *Curr. Opin. Plant Biol.* **10**, 453–460 (2007).
41. Yoon, G. M. & Kieber, J. J. 1-Aminocyclopropane-1-carboxylic acid as a signalling molecule in plants. *AoB PLANTS* **5**, plt017 (2013).
42. Li, Y. et al. Combined analysis of the fruit metabolome and transcriptome reveals candidate genes involved in flavonoid biosynthesis in *Actinidia arguta*. *Int. J. Mol. Sci.* **19**, 1471 (2018).
43. Ehlting, J. et al. Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* **19**, 9–20 (1999).
44. Ma, M. et al. Methyl jasmonate alleviates chilling injury by regulating membrane lipid composition in green bell pepper. *Sci. Hortic.* **266**, 109308 (2020).
45. Fang, H. et al. Stipe wall extension of Flammulina velutipes could be induced by an expansin-like protein from Helix aspersa. *Fungal Biol.* **118**, 1–11 (2014).
46. McQueen-Mason, S. & Cosgrove, D. J. Disruption of hydrogen-bonding between plant cell wall polymers by proteins that induce wall extension. *Proc. Natl Acad. Sci. USA* **91**, 6574–6578 (1994).
47. Bleecker, A. B. & Schaller, G. E. *Plant Physi*ol. **111**, 653–660 (1996).
48. Huang, F. C., Molnár, P. & Schwab, W. Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *J. Exp. Bot.* **60**, 3011–3022 (2009).
49. Wu, S. et al. Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the US National Plant Germplasm System watermelon collection. *Plant Biotechnol. J.* **17**, 2246–2258 (2019).
50. Urasaki, N. et al. Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* **24**, 51–58 (2017).
51. Sun, H. et al. Karyotype stability and unbiased fractionation in the paleo-allotetraploid cucurbita genomes. *Mol. Plant* **10**, 1293–1306 (2017).
52. Wang, J. et al. An overlooked paleotetraploidization in cucurbitaceae. *Mol. Biol. Evol.* **35**, 16–26 (2018).
53. Barrera-Redondo, J. et al. The genome of *Cucurbita argyrosperma* (silver-seed gourd) reveals faster rates of protein-coding gene and long noncoding rna turnoverand neofunctionalization within cucurbita. *Mol. Plant* **12**, 506–520 (2019).
54. Meguro, M. et al. Identification of the 7-hydroxymethyl chlorophyll a reductase of the chlorophyll cycle in arabidopsis. *Plant Cell* **23**, 3442–3453 (2011).
55. Diretto, G. et al. Silencing of beta-carotene hydroxylase increases total carotenoid and beta-carotene levels in potato tubers. *BMC Plant Biol.* **7**, 11 (2007).
56. Mosaab, Y. et al. Isolation and functional characterization of carotenoid cleavage dioxygenase-1 from *Laurus nobilis* L. (Bay Laurel) fruits. *J. Agric. Food Chem.* **63**, 8275–8282 (2015).
57. Rodrigo, M. J., Alquezar, B. & Zacarías, L. Cloning and characterization of two 9-cis-epoxycarotenoid dioxygenase genes, differentially regulated during fruit maturation and under stress conditions, from orange (*Citrus sinensis* L. Osbeck). *J. Exp. Bot.* **57**, 633–643 (2006).
58. Marin, E. et al. Molecular identification of zeaxanthin epoxidase of Nicotiana plumbaginifolia, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana*. *EMBO J.* **15**, 2331–2342 (1996).
59. Hothorn, M., Dabi, T. & Chory, J. Structural basis for cytokinin recognition by *Arabidopsis thaliana* histidine kinase 4. *Nat. Chem. Biol.* **7**, 766–768 (2011).
60. Gu, Y. et al. Identification fruit of a cellulose synthase-associated protein required for cellulose biosynthesis. *Proc. Natl Acad. Sci. USA* **107**, 12866–12871 (2010).
61. Chauvin, A., Lenglet, A., Wolfender, J. L. & Farmer, E. E. Paired hierarchical Organization of 13-lipoxygenases in arabidopsis. *Plants* **5**, 16 (2016).
62. Xu, D., Hu, M., Wang, Y. & Cui, Y. Antioxidant activities of quercetin and its complexes for medicinal application. *Molecules* **24**, 1123 (2019).
63. Chen, Y., Zheng, R., Jia, Z. & Ju, Y. Flavonoids as superoxide scavengers and antioxidants. *Free Radic. Biol. Med.* **9**, 19–21 (1990).
64. Hanano, S. et al. A systematic survey in *Arabidopsis thaliana* of transcription factors that modulate circadian parameters. *BMC Genomics* **9**, 182–182 (2008).
65. Koren, S., Walenz, B. P., Berlin, K., Miller, J. R. & Phillippy, A. M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

66. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–764 (2017).
67. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
68. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics*. Preprint at *arXiv* https://arxiv.org/abs/1303.3997 (2013).
69. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
70. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
71. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, I351–I358 (2005).
72. Hoede, C. et al. PASTEC: an automatic transposable element classification tool. *PLoS ONE* **9**, e91929 (2014).
73. Jurka, J. et al. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
74. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* **Chapter 4**, Unit 4.10 (2009).
75. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
76. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, II215–II225 (2003).
77. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
78. Blanco, E., Parra, G. & Guigo, R. Using geneid to identify genes. *Curr. Protoc. Bioinf.* **Chapter 4**, Unit 4.3 (2007).
79. Korf, I. Gene finding in novel genomes. *BMC Bioinf.* **5**, 59 (2004).
80. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* **19**, 189 (2018).
81. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–U121 (2015).
82. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
83. Tang, S., Lomsadze, A. & Borodovsky, M. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78 (2015).
84. Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M. & Buell, C. R. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* **7**, 327 (2006).
85. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
86. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
87. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
88. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
89. Marchler-Bauer, A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).
90. Koonin, E. V. et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
91. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
92. Boeckmann, B. et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
93. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
94. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
95. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
96. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16**, 284–287 (2012).
97. Lam-Tung, N., Schmidt, H. A., von Haeseler, A. & Bui Quang, M. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
98. Yang, Z. H. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
99. Han, M. V., Thomas, G. W. C., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
100. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
101. Tang, H. et al. jcvi: JCVI utility libraries. *Zenodo* https://doi.org/10.5281/zenodo.31631. (2015).
102. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
103. Zwaenepoel, A. & Van de Peer, Y. wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **35**, 2153–2155 (2019).
104. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, 593–601 (2014).
105. Michael, L. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550–550 (2014).
106. Fraga, C. G. et al. Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-mass spectrometry, XCMS, and chemometrics. *Anal. Chem.* **82**, 4165–4173 (2010).
107. Garcia, A., Barbas, C., Garcia, A. & Barbas, C. Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. Metabolics profiling. *Methods Mol. Biol.* **708**, 191–204 (2011).
108. Chen, Y. et al. RRLC-MS/MS-based metabonomics combined with in-depth analysis of metabolic correlation network: finding potential biomarkers for breast cancer. *Analyst* **134**, 2003–2011 (2009).
109. Chen, W. et al. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* **6**, 1769–1780 (2013).
110. Chong, J. & Xia, J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics* **34**, 4313–4314 (2018).
111. Thévenot, E. et al. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).
112. Jozefczuk, S. et al. Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* **6**, 364–364 (2010).