

Telomere-to-telomere genome assembly and comprehensive mutation library facilitate bitter gourd breeding and functional genomics research

Yilin Zhang^{2,5,10,11,12,*}, Xiaoyi Wang^{1,3,4,12}, Yan Wang^{2,12}, Yu Niu^{1,8,9}, Xu Han^{1,8,9}, Jingsheng Tan², Xinyu Ma², Qingzheng Zhang², Jie Liu^{1,6}, Zhijie Zhang^{1,6}, Wenying Gao^{1,6}, Wenjie Liang², Luyao Li^{1,7}, ZhaoHua Liu^{1,8,9}, Xingping Zhang^{2,*}, Hang He^{2,10,*}, Yan Yang^{1,4,8,9,*} and Renbo Yu^{1,4,8,9,*}

¹Tropical Crops Genetic Resources Institute, State Key Laboratory of Tropical Crop Breeding, Chinese Academy of Tropical Agricultural Sciences, Haikou/Sanya, Hainan 571101/572024, China

²National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang 261000, China

³Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, China

⁴Sanya Research Institute, Chinese Academy of Tropical Agricultural Sciences, Sanya, 572024, Hainan, China

⁵Laboratory of Plant and Environment Interaction Study, Tsientang Institute for Advanced Study, Hangzhou 310024, China

⁶College of Tropical Agricultural and Forestry, Hainan University, Haikou 570228, Hainan, China

⁷College of Horticulture & Forestry Sciences of Huazhong Agricultural University, Wuhan 430070, China

⁸Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture and Rural Affairs, Haikou 571101, China

⁹Key Laboratory of Tropical Crops Germplasm Resources Genetic Improvement and Innovation of Hainan Province, Haikou 571101, China

¹⁰State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agricultural Sciences and School of Life Sciences, Peking University, Beijing 100871, China

¹¹College of Life Sciences, Zhejiang University, Hangzhou 310058, China

¹²These authors contributed equally

*Correspondence: Yilin Zhang (yilinzhang@stu.pku.edu.cn), Xingping Zhang (xingping.zhang@pku-iaas.edu.cn), Hang He (hehang@pku.edu.cn), Yan Yang (yanyang@catas.cn), Renbo Yu (yurenbo612@163.com)

<https://doi.org/10.1016/j.xplc.2026.101779>

ABSTRACT

Bitter gourd (*Momordica charantia*) is an economically important vegetable crop with notable medicinal properties, yet its genetic improvement has been hindered by limited genomic resources and insufficient functional genomics tools. Here, we present a comprehensive genomic and functional resource platform to address these limitations. Using PacBio HiFi and Oxford Nanopore ultra-long sequencing technologies combined with Hi-C, we generated a telomere-to-telomere (T2T), gap-free genome assembly of the elite bitter gourd cultivar Y52, comprising 11 chromosomes with a final assembly size of 298.0 Mb. This high-quality assembly enabled precise annotation of complex genomic regions, including centromeres characterized through chromatin immunoprecipitation sequencing. We further constructed a large-scale ethyl methanesulfonate (EMS)-induced mutant library consisting of 3223 M₁ plants and performed whole-genome resequencing of 320 mutants, identifying approximately 76 002 induced mutations affecting 13 984 genes. Using this mutant population, we rapidly identified the causal gene *McEGY1*, responsible for chlorotic leaf and fruit phenotypes, through a MutMap-based approach. In addition, we conducted bulked segregant analysis and bulked segregant RNA sequencing to identify genomic regions and candidate genes associated with key agronomic traits, including fruit length, fruit tubercle formation, and powdery mildew resistance. Moreover, integrated transcriptomic and metabolomic analyses across multiple tissues and developmental stages generated comprehensive profiles of gene expression and metabolite accumulation, revealing tissue-specific regulatory networks. Finally, we integrated the genomic, mutant, transcriptomic, and metabolomic resources generated in this study to establish Bittergourd DB, a comprehensive platform supporting bitter gourd functional genomics and molecular breeding. Collectively, this study provides an unprecedented genomic and functional resource for bitter gourd, including a complete

T2T reference genome, an extensive EMS mutant library, trait-associated loci, and multi-omics maps, which will substantially accelerate functional genomics research and molecular breeding in this species.

Key words: T2T genome, centromere evolution, EMS mutant library, MutMap, QTL-seq, pan-transcriptomics

Zhang Y., Wang X., Wang Y., Niu Y., Han X., Tan J., Ma X., Zhang Q., Liu J., Zhang Z., Gao W., Liang W., Li L., Liu Z., Zhang X., He H., Yang Y., and Yu R. (2026). Telomere-to-telomere genome assembly and comprehensive mutation library facilitate bitter melon breeding and functional genomics research. *Plant Comm.* **7**, 101779.

INTRODUCTION

Bitter melon (*Momordica charantia*, $2n = 22$)—also known as bitter melon, bitter cucumber, and balsam pear—is a member of the Cucurbitaceae family and is widely cultivated across Asia and Africa (Cui et al., 2020). As an important vegetable crop, it exhibits numerous pharmacological properties, including effects on HIV infection and diabetes management; it thus plays important roles in human health and daily life (Lee-Huang et al., 1990). Under conditions of limited arable land and adverse environmental factors, breeding programs increasingly focus on improving crop yield. Bitter melon production faces the challenge of achieving higher yields with reduced land resources, an issue that can be effectively addressed through biotechnological approaches. For many years, basic research on bitter melon has primarily concentrated on the development of recombinant inbred lines and construction of genetic linkage maps via molecular markers (Cui et al., 2018, 2020; Gangadhara Rao et al., 2018; Sarao, 2021; Yang et al., 2022; Zhong et al., 2022). In contrast, research on functional genes in bitter melon has progressed considerably more slowly compared with other crops. In addition to the limited availability of elite germplasm, studies of functional gene discovery and genetic transformation in bitter melon remain underdeveloped relative to other crops, and very few gene resources are available for molecular breeding applications. These limitations constitute severe constraints on genetic improvement efforts in bitter melon. To overcome breeding bottlenecks and improve breeding efficiency, there is an urgent need to develop innovative germplasm resources and identify key genes that control important agronomic traits.

Ethyl methanesulfonate (EMS) is an effective mutagen that introduces random base-pair substitutions (primarily G:C to A:T transitions) in the genome during replication while causing minimal damage to plants (Wang et al., 2023). In *Arabidopsis thaliana*, approximately 700 mutations are induced per EMS mutant line (Jander et al., 2003). In *Citrullus lanatus*, another cucurbit species, an average of 401 mutations was identified per mutant individual in the M_2 generation (Deng et al., 2022). Due to advantages such as ease of control, high mutation rates, and minimal damage to plants, EMS mutagenesis has been widely applied in diverse crops, including rice (Henry et al., 2014), soybean (Li et al., 2017), maize (Lu et al., 2018), cabbage (Sun et al., 2022), cucumber (Zhang et al., 2022), and watermelon (Deng et al., 2022), thereby substantially advancing functional genomics research in these crops. Additionally, mutated genes in EMS populations can be efficiently identified through MutMap analysis (Abe et al., 2012). However, given the limited number of studies concerning EMS treatment in bitter melon, there is an urgent need to conduct systematic EMS mutagenesis experiments for this crop.

Despite some progress in the development of genetic and genomic resources for bitter melon, additional efforts are needed to establish a more integrated multi-omics platform that can accelerate genetic research and support trait improvement. A complete reference genome represents a fundamental resource for advancing genetic studies and plays a critical role in functional genomics research (Garg et al., 2024). In recent years, genomic resources for bitter melon have substantially improved. The first bitter melon reference genome (OHB3-1) was published in 2017 using Illumina short-read sequencing (Urasaki et al., 2017); two additional draft genomes (Dali-11 and Goya) were reported in 2020 (Cui et al., 2020; Matsumura et al., 2020). Along with these draft assemblies, the first T2T genome of the bitter melon variety Mca was generated in 2023 using PacBio high-fidelity (HiFi) sequencing and Hi-C technologies (Fu et al., 2023). Nevertheless, these assemblies contain sequence gaps within highly repetitive regions, such as telomeres and centromeres; for example, the Mca genome retains three gaps and lacks four telomeric sequences. Therefore, the establishment of more complete reference genomes for elite inbred cultivated bitter melon lines remains essential for advancing genetic and genomic studies.

Centromeres are essential chromosomal regions that ensure accurate chromosome segregation during cell division; however, their complex repetitive structure has made them difficult to assemble and characterize (Włodzimierz et al., 2023). Chromatin immunoprecipitation sequencing (ChIP-seq) targeting centromere-specific histone variants offers a powerful strategy for identifying and analyzing centromeric regions (Naish et al., 2021). In bitter melon, as in many other crops, centromere organization and evolution remain largely unexplored, which limits understanding of genome structure and chromosome stability. Characterization of these regions would provide insights into bitter melon genome evolution and may identify features that could be exploited in breeding programs.

Integrative multi-omics approaches have emerged as powerful tools for dissecting complex biological systems and identifying key regulators of important traits (Misra et al., 2019). By integrating transcriptomic and metabolomic analyses across diverse tissues and developmental stages, researchers can establish comprehensive catalogs of genes and metabolites that coordinate developmental processes and stress responses (Han et al., 2023). Such multi-dimensional datasets are particularly valuable for medicinal plants such as bitter melon, in which specialized metabolites substantially contribute to therapeutic properties (Behera et al., 2010; Sur and Ray, 2020). Tissue-specific gene expression patterns and metabolite accumulation profiles can reveal regulatory networks that govern important traits, including fruit development, nutritional quality, and biosynthesis of medicinal compounds (Li et al., 2024).

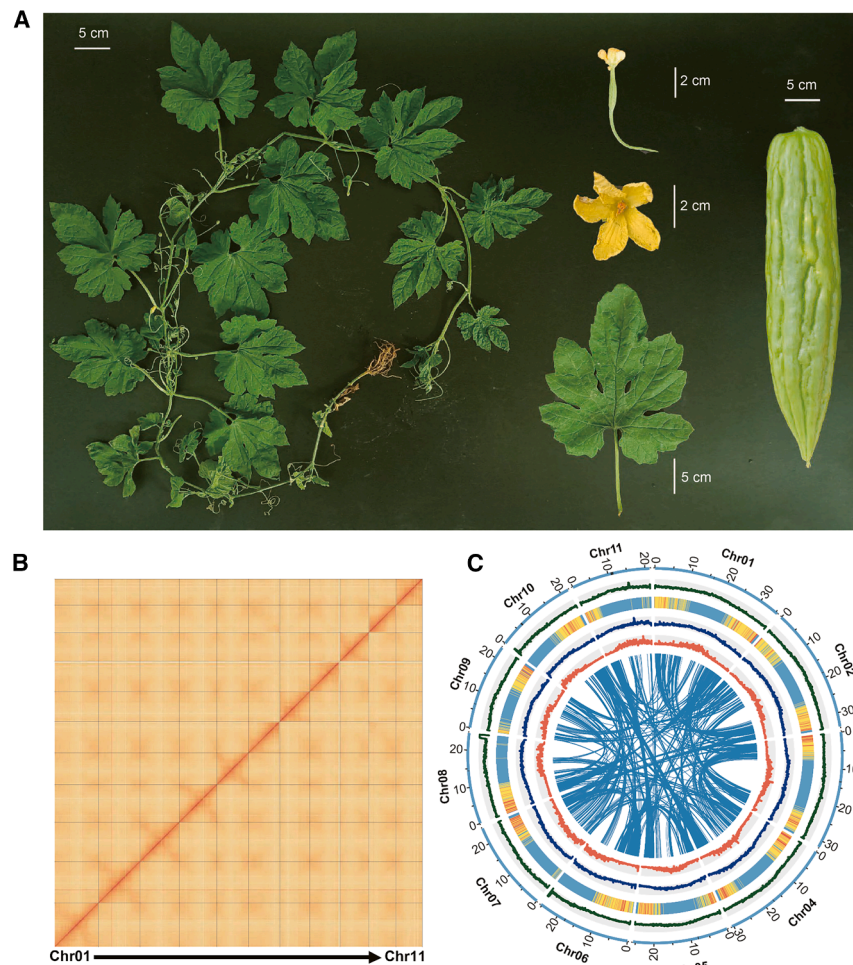


Figure 1. Overview of the gap-free reference genome of bitter gourd Y52.

(A) Representative morphological characteristics of the bitter gourd accession Y52, including the whole plant (left), male flower (top middle), female flower (center middle), mature leaf (bottom middle), and mature fruit (right). Scale bars are indicated.

(B) Hi-C interaction heatmap of the bitter gourd genome, displaying contact frequencies among chromosomes (Chr01–Chr11).

(C) Circos plot of the bitter gourd genome showing the 11 chromosomes. From outer to inner tracks: chromosome ideograms with scale (Mb), GC content, gene density, Gypsy density, Copia density, and syntenic relationships among chromosomal regions, represented by blue links in the center.

500 to 650 g (Figure 1A). Additionally, Y52 exhibits a degree of disease resistance, including resistance to *Fusarium* wilt and powdery mildew. Thus, we selected the elite inbred line Y52 for T2T gap-free genome assembly, and multiple sequencing platforms were used to generate genomic data (Supplemental Table 1). Genome survey analysis indicated that the estimated genome size of Y52 was ~300 Mb, with a heterozygosity ratio of 0.0758% (Supplemental Table 2). We generated 17.83 Gb (~59.85×) of HiFi reads and 61.44 Gb (~206.16×) of Oxford Nanopore Technology (ONT) ultra-long reads using the PacBio Sequel II and MinION platforms, respectively. The N50 lengths of the

HiFi and ONT reads were 15.8 kb and 94.03 kb, respectively. Additionally, we generated 19.99 Gb (~67.09×) of Hi-C data to anchor contigs.

Draft genome assemblies generated from PacBio HiFi reads contained only contigs and achieved a contig N50 length of 25.89 Mb (Table 1). After removal of contigs smaller than 500 kb, the resulting genome assembly consisted of 11 chromosomes, including one gap and 20 telomeres. Subsequently, a gap-free assembly was obtained by integrating the HiFi-based hifiasm assembly with supplementary assemblies from NextDenovo and ultra-long nanopore reads (Supplemental Figures 1A–1F). Ultimately, the integrated hifiasm assembly produced a T2T reference genome for bitter gourd comprising 11 chromosomes with a total length of 298.0 Mb (Figure 1B). Using the canonical seven-base telomeric repeat sequences (5'-CCCATTT and 3'-TTTAGGG) as queries, we identified 22 telomeres, which enabled the construction of 11 fully resolved T2T pseudomolecules in the Y52 genome (Supplemental Table 3).

Our annotation pipeline comprised three integrated phases. Initially, repetitive genomic elements were identified and annotated as transposable elements (TEs); these sequences were subsequently masked to facilitate accurate gene prediction. In total, repetitive sequences constituted approximately 136.94 Mb,

In this study, we aimed to establish a comprehensive research platform to support bitter gourd research and molecular breeding using the elite inbred line Y52. First, we assembled a telomere-to-telomere (T2T) complete genome of bitter gourd Y52 and then developed a large-scale EMS-induced mutant population and sequenced 320 mutant lines. Using these genetic resources, we identified causal gene mutations in *EGY1* Y52 mutants through a MutMap-based approach. We also generated tissue-specific transcriptome and metabolome maps to characterize gene expression patterns and metabolite profiles across various tissues. Furthermore, we utilized bulked segregant analysis (BSA) and bulked segregant RNA sequencing (BSR) to identify genomic regions associated with important agronomic traits, such as fruit length and disease resistance. This study provides high-quality genomic resources, candidate loci for several key bitter gourd traits, an indexed mutant library, and comprehensive gene and metabolite expression atlases to support functional genomics research and genetic improvement of bitter gourd.

RESULTS

A T2T reference genome assembly for bitter gourd

Y52 is an excellent breeding parent due to several desirable traits, including a female flower rate of 60%, green-skinned fruit measuring 32–38 cm in length with a transverse stem diameter of approximately 7.5 cm, and individual fruit weights ranging from

Genome assembly	total length of assembly (Mb)	298.0 (295.6)
	number of chromosomes	11
	number of gaps	0 (3)
	Contigs N50 (Mb)	25.89 (25.38)
	number of telomeres	22 (18)
	number of centromeres	11
	GC content	35.64% (35.92%)
Genome evaluation	PacBio HiFi reads mapping rate	99.79%
	ONT reads mapping rate	99.59%
	Hi-C reads mapping rate	99.71%
	completeness (BUSCO)	98.50% (98.02%)
	LTR assembly index (LAI)	25.87 (13.40)
	quality value (QV)	55.25 (45.20)
Repetitive elements (Mb)	retrotransposons size	79.58
	DNA transposons size	10.48
	other repetitive elements	46.88
	total size of repetitive elements	136.94
Annotated genes	protein-coding genes	25,189
	mean gene length (bp)	2908.39
	genes with eggNOG annotation	23,994
	completeness (BUSCO)	98.3% (96.6%)

Table 1. Summary of the Y52 genome assembly and annotation

The numbers in brackets in the third column are the indicators of the published Mca genomes (Fu et al., 2023).

representing 45.96% of the assembled Y52 genome (Supplemental Table 4). Regions with high gene density exhibited lower TE density (Supplemental Figure 2A), consistent with previously reported distribution patterns (Qin et al., 2021; Deng et al., 2022). TE insertion was associated with increased gene expression levels (Supplemental Figure 2B), suggesting that TEs enhance transcriptional activity in certain contexts by modifying regulatory elements. However, TE insertion also caused complete silencing of 4304 genes (a high proportion among TE insertion categories), likely due to structural disruption (Supplemental Figure 2C). Non-TE genes exhibited stronger tissue-specific expression, whereas genes with TE insertions showed more prominent constitutive expression (Supplemental Figure 2D), indicating that TE-mediated transcriptional enhancement is a broad, non-specific mechanism that promotes persistently high expression across different tissues. After repeat masking, gene-structure prediction was conducted on the masked genome. Predictions derived from these complementary approaches were subsequently integrated using EvidenceModeler, which refined gene models and excluded those with lower confidence. Ultimately, we identified a total of 25 189 high-confidence protein-coding gene models in the Y52 genome assembly (Table 1). Additionally, homology-based annotation of non-coding RNAs identified a diverse set of RNA genes, including 1130 transfer RNA genes, 1327 ribosomal RNA genes, and 4057 small nucleolar RNA (snoRNA) genes (Supplemental Table 5). Y52 contains a more complete set of 45S rDNA than Mca: 118 copies of 45S rDNA in Y52 compared with only 14 copies in Mca (Supplemental Table 6). To further characterize genomic architecture and variation, we examined GC content, gene density, Gypsy density,

and Copia density across the 11 chromosomes with sliding windows of 500 kb (Figure 1C).

The quality and completeness of the T2T genome assembly of Y52 were rigorously evaluated using multiple complementary approaches. First, collinearity analysis demonstrated strong genomic synteny between Y52 and the reference genome of Mca, further supporting the reliability of the assembly. Second, resequencing data showed an average mapping rate of 99.51% and genome coverage of 99.90% when aligned to the Y52 T2T reference genome (Supplemental Table 7), indicating high assembly accuracy and completeness at the nucleotide level. Third, benchmarking universal single-copy orthologs (BUSCO) analysis indicated that 98.50% (1590 of 1614 embryophyte BUSCO genes) of conserved single-copy genes were fully assembled and intact (Supplemental Table 8). Moreover, the genome assembly exhibited a consensus quality value (QV) score of 55.25, with individual chromosomes ranging from 40.50 to 68.92, reflecting high nucleotide accuracy across the genome (Supplemental Table 9). Finally, the long terminal repeat assembly index (LAI) was calculated as 25.87 for the Y52 genome (Table 1). Collectively, these assessments confirm that the newly generated T2T assembly of Y52 achieves high degrees of accuracy, completeness, and structural integrity.

The Cucurbitaceae landscape of centromeres

Through integrated analysis of McCENH3 ChIP-seq and Hi-C chromatin interaction data, we mapped centromeric regions across all 11 chromosomes; each chromosome contained a

single distinct CENH3-binding domain. The combined length of these centromeric regions was 8.97 Mb, with considerable size variation ranging from 0.58 Mb (Chr06) to 2.00 Mb (Chr02), and a mean length of 1.32 Mb (Supplemental Table 10).

Detailed sequence analysis demonstrated that Y52 centromeres predominantly consist of TE sequences; long terminal repeat (LTR)/Gypsy retrotransposon fragments represent the most abundant class (45.42%), followed by LTR/Copia fragments (13.47%) and DNA transposons (14.37%) (Figure 2; Supplemental Figures 3A and 4). To distinguish between intraspecific and interspecific variation, we conducted a comprehensive centromere analysis using a second bitter gourd accession, Mca (Fu et al., 2023). Because the published Mca genome lacks CENH3 ChIP-seq data, we utilized an integrative strategy that combined peaks obtained by mapping Y52 CENH3 ChIP-seq reads to the Mca genome, TE distribution patterns, and synteny with the Y52 genome. This multi-evidence approach allowed reliable definition and refinement of centromere boundaries in Mca (Supplemental Figure 5). Our analysis indicated that, similar to Y52, Mca centromeres are entirely dominated by TEs and lack satellite sequences, confirming that TE-based centromeres represent a stable, lineage-specific feature of bitter gourd. However, the proportions of Gypsy elements and DNA transposons in Mca are substantially higher than those in Y52; the relative abundances of other TE types are correspondingly lower (Supplemental Figure 5). These results suggest that although the absence of satellites and the TE-dominated architecture constitute conserved lineage-specific features, variation in specific TE composition reflects intraspecific polymorphism. Such variation is likely driven by ongoing TE dynamics and lineage-specific transposition bursts, which may be associated with divergence among *M. charantia* varieties.

To investigate centromere evolution in Cucurbitaceae, we analyzed the centromeric composition of cucumber (*Cucumis sativus*), watermelon (*Citrullus lanatus*), melon (*Cucumis melo*), and pumpkin (*Cucurbita maxima*) (Deng et al., 2022; Mo et al., 2024; Zeng et al., 2024; Tian et al., 2025). Cucumber centromeres are primarily organized as arrays of 177-bp satellite repeat units (59.89%) interspersed with intact LTR retrotransposon insertions (15.13%) (Figure 2; Supplemental Figures 3B and 6). Of particular interest, our analysis clarified the origin of centromeric satellite repeats in cucumber. We found that a CsMITE sequence shares 85.31% sequence similarity with most 177-bp satellite monomers (Supplemental Figure 7A). These findings suggest that the 177-bp satellite monomer sequences may have originated from ancestral miniature inverted-repeat transposable element (MITE) insertions into centromeric regions during evolution. Centromere analysis in watermelon revealed structural features more similar to those of bitter gourd than cucumber; sequences were dominated by LTR elements (Gypsy fragments: 52.16%; Copia fragments: 14.73%) and lacked canonical satellite repeats (Figure 2; Supplemental Figures 3C and 8). The centromeres of pumpkin comprise six satellite monomers, and their TE composition is dominated by Gypsy elements (Figure 2; Supplemental Figures 3D and 9). In melon, centromeres are primarily occupied by a 354-bp satellite repeat, with Gypsy and Copia retrotransposons inserted at various proportions across different chromosomes (Figure 2; Supplemental Figures 3E and 9). Phylogenetic analysis of satellite monomers across these species revealed that pumpkin satellites form two independent clades, with CEN253 evolving independently.

Importantly, the satellite sequences exhibit species-specific evolution without cross-species homology, confirming the independent evolutionary origins of satellite repeats among Cucurbitaceae lineages (Supplemental Figure 7B). By integrating centromere analyses across these five Cucurbitaceae species, we identified three distinct compositional types: TE-dominated (bitter gourd and watermelon), satellite-dominated (cucumber), and co-existing TE-dominated and satellite-dominated (melon and pumpkin). The evolutionary model of Cucurbitaceae centromeres therefore implies repeated structural transitions between satellite-dominated and TE-dominated states (Figure 2). Comparison with the phylogenetic relationships among Cucurbitaceae species (Guo et al., 2020) indicated that centromere composition does not correspond with species evolutionary relationships. Instead, it reflects dynamic and largely independent evolutionary trajectories of this genomic region across lineages. Such diversity is consistent with observations in Solanaceae (Chen et al., 2024) and eukaryotic centromeres overall (Villasante et al., 2007), suggesting that centromeres might repeatedly “re-invent” their structural organization across different plant families.

Construction of the Y52 EMS mutant library

To prevent seed damage and germination inhibition caused by high EMS concentrations and prolonged exposure, we initially determined optimal mutagenesis conditions in bitter gourd. In the preliminary assay, the germination rate of the control group (treated with ddH₂O) exceeded 90%. Seeds treated with 0.5% EMS for either 4 h or 10 h exhibited germination rates below 50%. In contrast, treatment with 0.2% EMS for 4 h produced limited mutagenic effects. Seeds treated with 0.2% EMS for 10 h showed a germination rate of approximately 50%–60%, which was close to the half-lethal dose. Seeds of the inbred line Y52 were used as the mutagenic material, and EMS served as the mutagen. In total, 3223 M₁ (first-generation mutant) plants were obtained (Figure 3A). In the M₂ generation, 241 families (7.5%) displayed visible phenotypic variations, which were primarily categorized as dwarf mutants, fruit mutants, leaf mutants, and plant architecture mutants. Among the variations, fruit-shape mutations were most abundant, with 99 mutants identified (Supplemental Table 11). These included 36 mutants with blunt bases, 7 with blunt ends, 12 with fruit bending, 16 spindle-shaped mutants, 3 color mutants, 12 fruit-wart mutants, and 1 twin-fruit mutant. Leaf mutations, another prominent variation, represented 65 mutants, including 29 yellow-leaf mutants, 24 light-green-leaf mutants, 5 rolled-leaf mutants, 2 albino-leaf mutants, 2 dried-leaf mutants, and 2 piebald-leaf mutants. Additionally, 49 dwarf mutants were identified, which could accelerate breeding programs aimed at developing dwarf plant types. Furthermore, 2 gynocious mutants and 10 sterility mutants were observed; line M617 produced 3 female flowers per node. Additional mutants included 4 tendril mutants, 7 meristem-defect mutants, 2 stem mutants, 1 lateral branch mutant, 1 stem villus mutant, and 1 senescence mutant (Supplemental Table 12). Given the scarcity of bitter gourd mutants, this mutant collection represents an important genetic resource and is expected to substantially enhance bitter gourd breeding efforts.

Whole-genome resequencing and mutation identification of 320 mutant lines

To systematically identify nucleotide mutations induced by EMS treatment, we performed whole-genome sequencing of 320

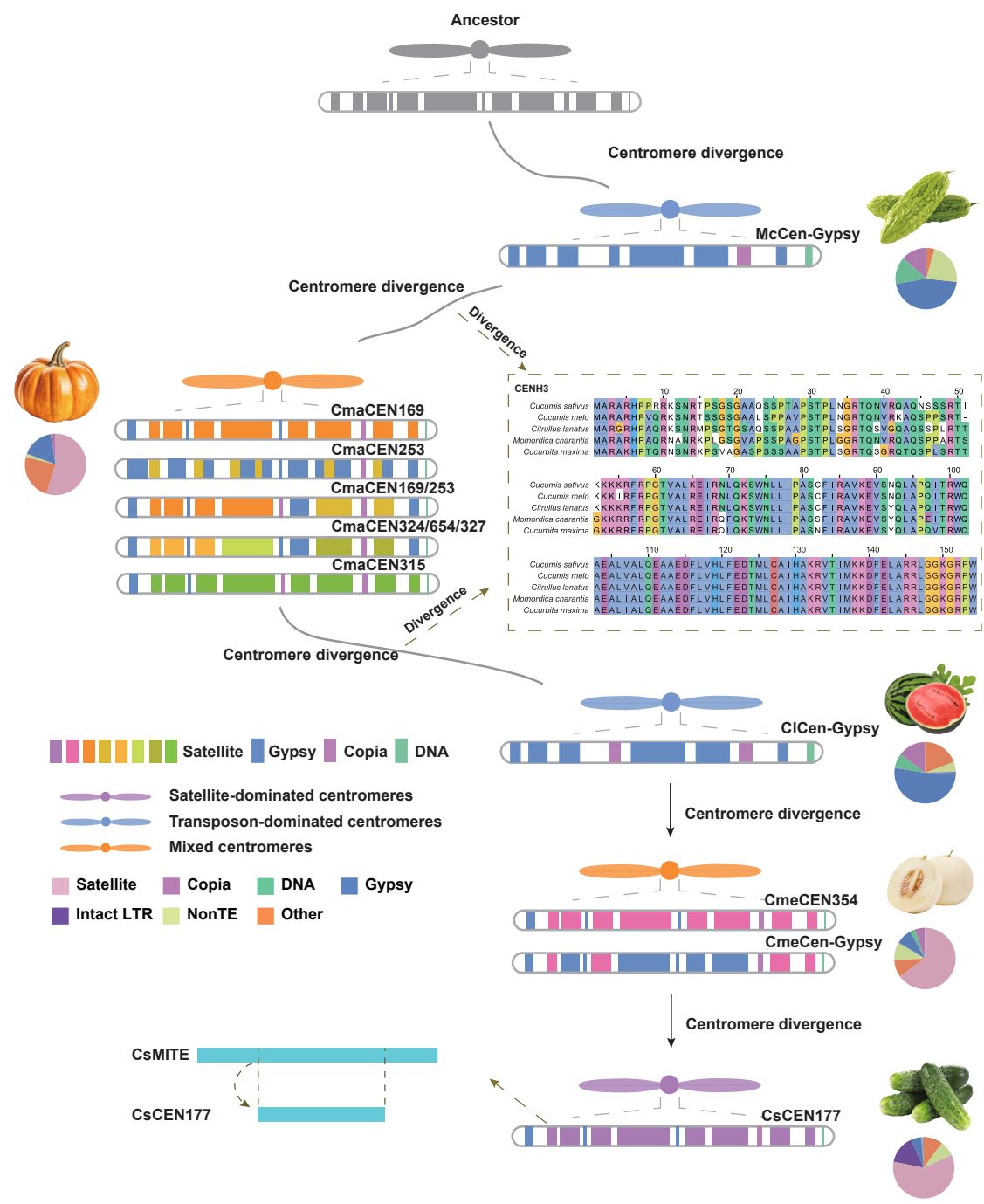


Figure 2. Evolutionary divergence and structural diversification of centromeres in Cucurbitaceae.

This schematic illustrates the dynamic evolutionary trajectory of centromere organization from a hypothetical common ancestor to lineage-specific architectures driven by the proliferation of transposable elements (TEs) and satellite repeats. The phylogeny reveals three distinct structural patterns: *Momordica charantia* and *Citrullus lanatus* retain TE-dominated centromeres enriched with Gypsy retrotransposons (blue segments), whereas *Cucurbita maxima* and *Cucumis melo* display mosaic structures comprising both TEs and diverse satellite arrays. *Cucumis sativus* represents a homogenized satellite-dominated configuration, in which the predominant CsCEN177 repeat likely originated from a CsMITE-like precursor (bottom-left inset). Despite the rapid turnover of centromeric DNA, the CENH3 protein sequence remains highly conserved across these lineages (middle alignment). Chromosome bars provide schematic representations of repeat organization (purple, Copia; blue, Gypsy; green, DNA transposons; other colors, species-specific satellites), and pie charts indicate the genomic abundance of these components.

mutant lines: 27 M₂ lines and 293 M₃ lines (Supplemental Figure 10A). In total, approximately 7.59 TB of clean sequencing data were generated and each sample yielded an average of 23.7 Gb, corresponding to an average sequencing

depth of 79.55×. Clean short reads were aligned to the gap-free Y52 reference genome, achieving average alignment rates and genome coverage of 99.51% and 99.90%, respectively (Supplemental Table 7). These results demonstrate the utility of

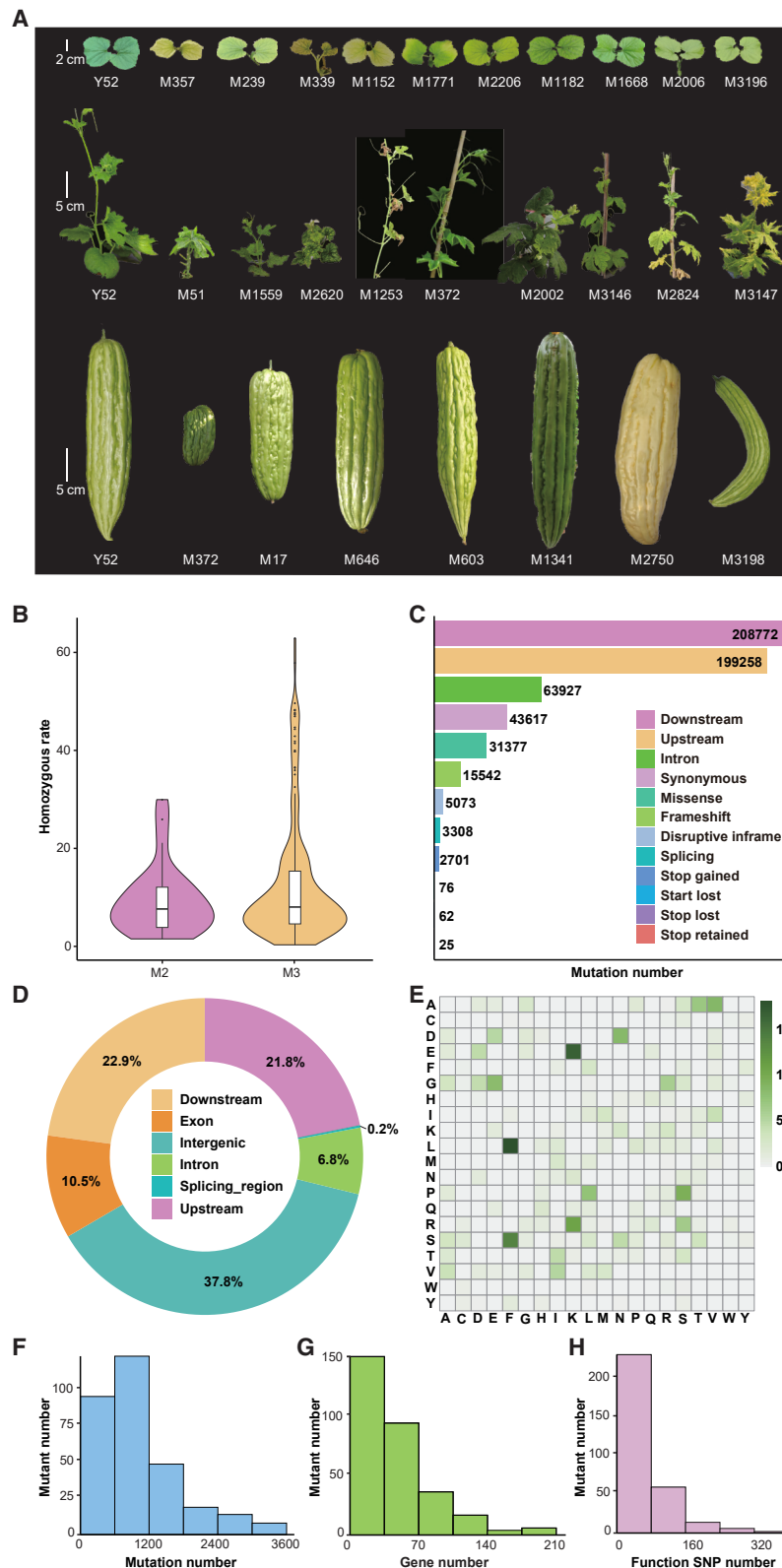


Figure 3. Phenotypes and characteristics of the bitter melon Y52 EMS mutant library.

(A) The first row depicts leaf-color mutants observed at the seedling stage. M339 and M357 are albino leaf mutants; M239, M1152, M1771, and M2206 are yellow leaf mutants; and M1182, M1668, M2006, and M3196 are light-green leaf mutants. In the second row, M51, M1559, and M2620 are dwarf mutants; M1253 is a dried-leaf mutant; M372 exhibits inwardly rolled leaves; M2002 and M3146 display downwardly rolled leaves; M2824 shows early senescence; and M3147 demonstrates abundant stem villus. The third row illustrates fruit mutants. M372 produces very short fruits measuring no more than 10 cm in length. The fruits of M17 and M646 display blunt ends. M603 shows a spindly fruit phenotype. M1341 exhibits stripe-like warts on the fruit peel. M2750 produces yellow fruit peel, whereas M3198 exhibits bent fruit.

(B) Homozygous mutation rates in the M_2 ($n = 27$) and M_3 ($n = 293$) generations.

(C) Detailed annotation of variant effects on genes.

(D) Genomic distribution of variant effects on genes.

(E) Amino acid substitutions detected in the mutants.

(F) Distribution of mutation numbers in each mutant line.

(G) Distribution of genes containing functional SNPs in each mutant.

(H) Distribution of functional SNP numbers in each mutant.

mutation calls and considerably improved identification accuracy (Deng et al., 2022).

We identified a total of 481 184 mutations across the genome, including 368 930 SNPs and 112 254 InDels. In the M_2 and M_3 generations, the proportion of homozygous mutations reached 9.79% and 12.87%, respectively, indicating progressive fixation of mutant genotypes (Figure 3B). Analysis of InDel mutations revealed lengths that predominantly ranged from 1 to 254 bp. The most frequent InDels were 1-bp insertions and deletions (29 425), followed by 2-bp insertions and deletions (14 856), 3-bp insertions and deletions (9284), and 4-bp insertions and deletions (8780) (Supplemental Table 13). These mutations were relatively evenly distributed across chromosomes, with an average mutation density of 161 mutation sites per 100 kb genomic region (Supplemental Figure 11). This distribution pattern provides a comprehensive overview of EMS-induced genetic variation across the genome and establishes a foundation for subsequent functional genomic analyses.

In addition to evaluating mutation distribution and frequency at the genome-wide level, we assessed the quality of the mutant library by

the T2T Y52 genome assembly as a reliable reference for mutation detection in the EMS mutant collection. Furthermore, the use of a T2T reference genome with the same genetic background as the mutants substantially reduced false-positive

examining the number and density of functionally effective mutations. SnpEff was used to predict the functional impact of each mutation (Figure 3C). Among all detected mutations, 29.40% were located within genic regions, affecting a total of 18 159 genes

(72.09% of all genes). According to the reference genome annotation, mutations most often occurred in intergenic regions (37.75%), followed by upstream (21.82%), downstream (22.87%), and intronic (6.77%) regions (Figure 3D). Regarding SNP mutations, C/G → T/A transitions predominated (59.77%), followed by T/A → C/G (13.68%), T/A → A/T (9.38%), C/G → A/T (5.74%), T/A → G/C (5.54%), and C/G → G/C (5.88%) (Supplemental Figure 10C and Supplemental Table 14). In terms of amino acid substitutions, the conversion from leucine (L) to phenylalanine (F) occurred most frequently, reaching 5.57%, which is considerably higher than the overall average amino acid substitution frequency of 0.25% (Figure 3E). Notably, the conversion from glutamic acid to lysine resulted from a single G>A nucleotide substitution, the most common nucleotide change detected among mutants. Additionally, substitutions involving leucine (5447; 8.35%), serine (6765; 10.37%), and arginine (4580; 7.02%) were observed at relatively high frequencies (Supplemental Table 15). The number of mutations per mutant widely varied (Figure 3F), ranging from 1923 (M2016-5) to 78 677 (M1349-9) SNPs and from 839 (M2673-1) to 9458 (M3090) InDels, with averages of 10 450 SNPs and 2882 InDels per plant (Supplemental Figure 10D). Within coding sequence (CDS) regions, we detected 76 002 SNPs affecting 13 984 genes (55.52% of the total gene set). Among these SNPs, 40 921 produced alterations in 11 503 protein sequences and thus were classified as functional SNPs. On average, each gene contained seven functional SNPs (Supplemental Figure 10E). Furthermore, each mutant harbored an average of 50 genes containing functional SNPs (Figure 3G). The mutant with the fewest functional SNP-containing genes was M3216-2, which harbored no genes, whereas mutant M1757-19 carried the highest number (275 genes). Additionally, each mutant contained approximately 69 functional SNPs on average (Figure 3H); missense variants represented the most abundant type, averaging approximately 82 per mutant (Supplemental Table 16). These mutation characteristics confirm the utility of our EMS mutagenesis protocol and provide critical parameters for predicting functional consequences of the observed genetic variations.

To further validate the accuracy of SNP mutations identified using Y52 as the reference genome, we randomly selected 20 SNPs representing diverse mutation types for verification across 20 mutants. The validation results demonstrated an accuracy rate of 95% for identified mutation sites, whereas non-mutated sites showed an accuracy rate of 5%. Additionally, we randomly selected 15 InDels (9 insertions and 6 deletions) for validation by Sanger sequencing across 15 mutants. Among these, 13 of the 15 tested loci were successfully validated, and two non-mutant loci were confirmed as correct (Supplemental Table 17). We also assessed the number of mutants required to construct a saturated bitter melon mutant library. Across the 320 analyzed mutants, we identified 40 921 functional SNP mutations; each mutant harbored an average of 69 functional SNPs. Based on the total number of bitter melon genes (25 189) and the average number of functional mutations per mutant (50), our calculations indicate that approximately 2319 mutants would be required to achieve 99% probability of covering all functional genes in the Y52 genome. The EMS mutant library generated for bitter melon Y52 produced more than 3000 M₁

seeds or plants, providing a mutation density sufficient to cover nearly all protein-coding genes.

In conclusion, the large-scale mutant library established in this study, combined with extensive sequencing of additional mutants, will substantially facilitate future research concerning the functional genomics of bitter melon.

Identification of the causal gene associated with the chlorotic phenotype

M2750 is a mutant characterized by yellow leaves and fruits, and this phenotype was confirmed in the M₃ generation (Figure 4A). We conducted a cross between M2750 and Y52, producing F₁ progeny that exhibited green leaves and fruits. In the F₂ generation, the yellow leaf and fruit phenotype reappeared at a frequency of approximately 25%. These results suggest that the chlorotic phenotype is controlled by a single recessive gene. We utilized the MutMap approach to identify the causal gene responsible for this phenotype (Figure 4B). Genome-wide analysis detected 63 mutations with an SNP index (Abe et al., 2012) greater than the 99% confidence threshold. Among these variants, six mutations were located within CDS regions. Three of these were synonymous mutations that did not alter amino acid sequences, whereas the remaining three were non-synonymous mutations that caused amino acid substitutions (Supplemental Table 18). One candidate SNP, Chr06_12144482 (G/A), was located in the second exon of gene Y52_Ch0601641. This mutation converted a glycine-encoding codon (GGG) to a glutamic acid-encoding codon (GAG), yielding a missense mutation (Figure 4C).

The homologous gene At5g35220 in *Arabidopsis thaliana* encodes a chloroplast precursor protein named ETHYLENE-DEPENDENT GRAVITROPISM-DEFICIENT AND YELLOW-GREEN 1 (EGY1) (Fu et al., 2025). The *Arabidopsis* *egy1* mutant exhibits a chlorotic phenotype similar to that observed in the bitter melon mutant, characterized by yellow leaves and fruits. The *AtEGY1* gene encodes a 59-kDa metalloprotease that is localized in the chloroplast and plays an essential role in chloroplast development, as evidenced by pale-green leaves observed in the *egy1* mutant. Additionally, the *CmEGY1* gene in melon and the *CsEGY1* gene in cucumber reportedly have critical roles in golden fruit formation. Therefore, we propose that gene Y52_Ch0601641 is a strong candidate for the source of the chlorotic phenotype in bitter melon, and we designated this gene as *McEGY1*.

In both the mutant parental DNA and the pooled F₂ sample, this locus was detected 149 times, and all sequencing reads carried the mutant A allele. In contrast, in the pooled WT DNA sample, the locus was detected 97 times; all reads corresponded to the WT G nucleotide (Supplemental Table 18). These results imply that the yellowing phenotype of the *egy1* mutant is controlled by a single recessive nuclear gene. To clarify whether the G/A SNP mutation in *McEGY1* was responsible for the observed chlorotic leaf and fruit phenotypes, we developed a cleaved amplified polymorphic sequence molecular marker to distinguish WT and chlorotic mutant seedlings in the F₂ population. Genotyping analysis clearly demonstrated that the G/A SNP in *McEGY1* co-segregated with the phenotype in the F₂

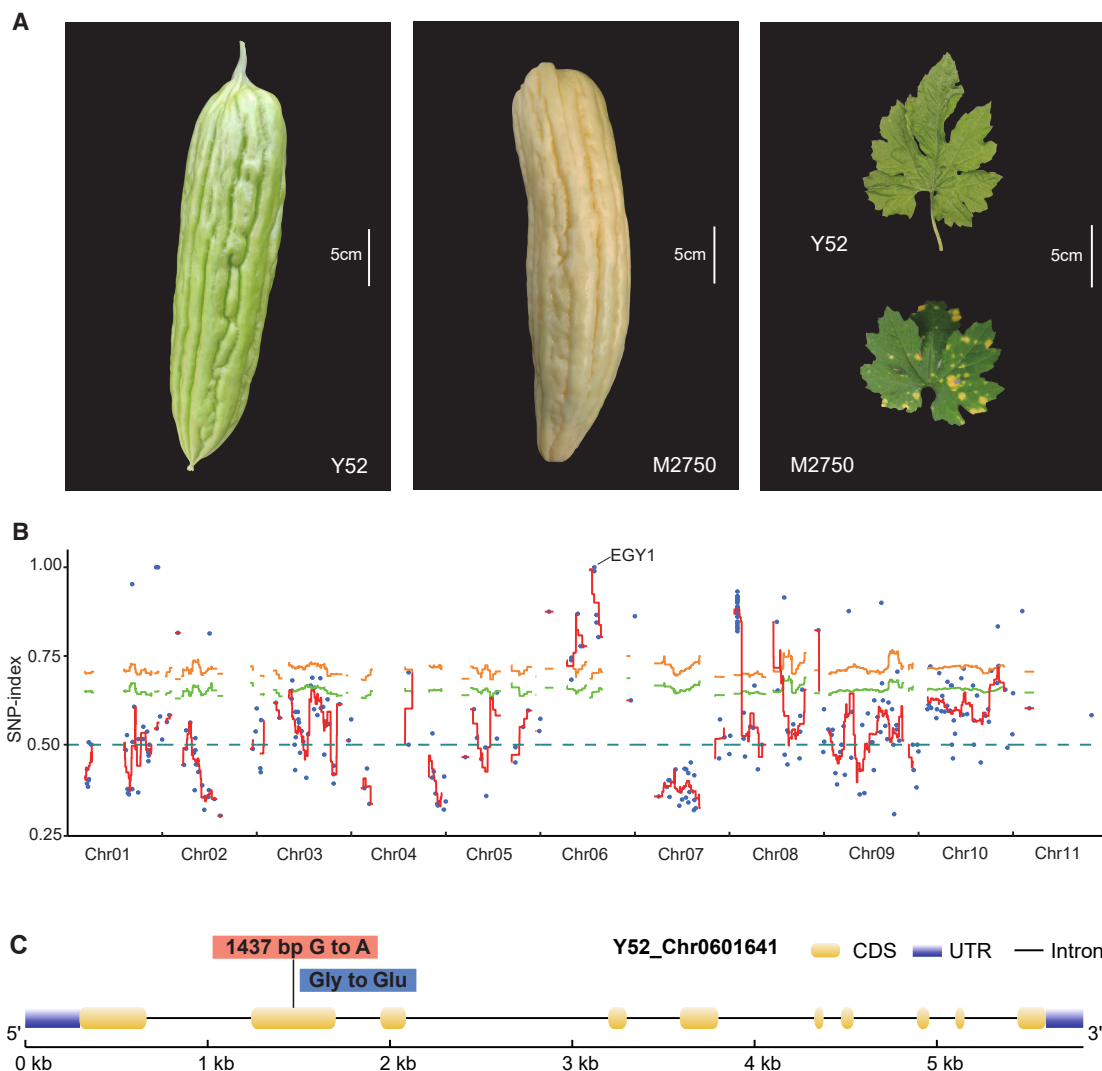


Figure 4. Phenotypic characterization and MutMap analysis of bitter melon mutant M2750.

(A) Comparison of fruit, leaf, and whole-plant phenotypes between the wild type (WT) and the M2750 mutant.

(B) SNP-index plot showing the distribution of allele frequencies in the F_2 segregating pool derived from mutant M2750 across the 11 chromosomes. Colored dots indicate SNP-index values calculated for individual SNP positions. The red line represents the average SNP-index values calculated using a sliding-window approach. Green and orange dashed lines indicate the statistical thresholds corresponding to the 95% and 99% confidence intervals, respectively. The candidate gene *EGY1* and the associated SNP on Chr06 are indicated.

(C) Gene structure and mutation location of the candidate gene Y52_Ch0601641 (*EGY1*). Yellow boxes represent coding sequences (CDS), blue boxes indicate untranslated regions (UTRs), and black lines denote introns. A G/A single-nucleotide substitution at position 1437 bp within the coding region results in an amino acid change from glycine (Gly) to glutamic acid (Glu).

segregating population. All individuals exhibiting chlorotic leaves or fruits were homozygous AA at this locus (Supplemental Figure 12A), whereas individuals with the WT phenotype were homozygous GG. Phylogenetic analysis of *EGY1* indicated that the gene originated from green algae and that its function has been conserved during evolution and within the Cucurbitaceae family (Supplemental Figures 12B–12D). Overall, these findings demonstrate that the EMS mutant population, combined with the T2T Y52 reference genome, enables rapid and accurate identification of functional genes in bitter melon.

Collectively, the above results indicate that the combination of a high-quality genome assembly and an EMS mutant population pro-

vides a valuable resource for functional gene analysis in bitter melon.

Expansion of bitter melon functional genomics through BSA reveals trait-associated genes

Although some progress has been made in mapping genes associated with specific traits in bitter melon, many genomic regions linked to important agronomic traits remain unidentified, which has hindered in-depth functional genomic research in this crop (Wang and Xiang, 2013; Rao et al., 2021; Yang et al., 2022; Karube et al., 2023; Lavale et al., 2023; Panigrahi et al., 2024). To address this knowledge gap, we conducted

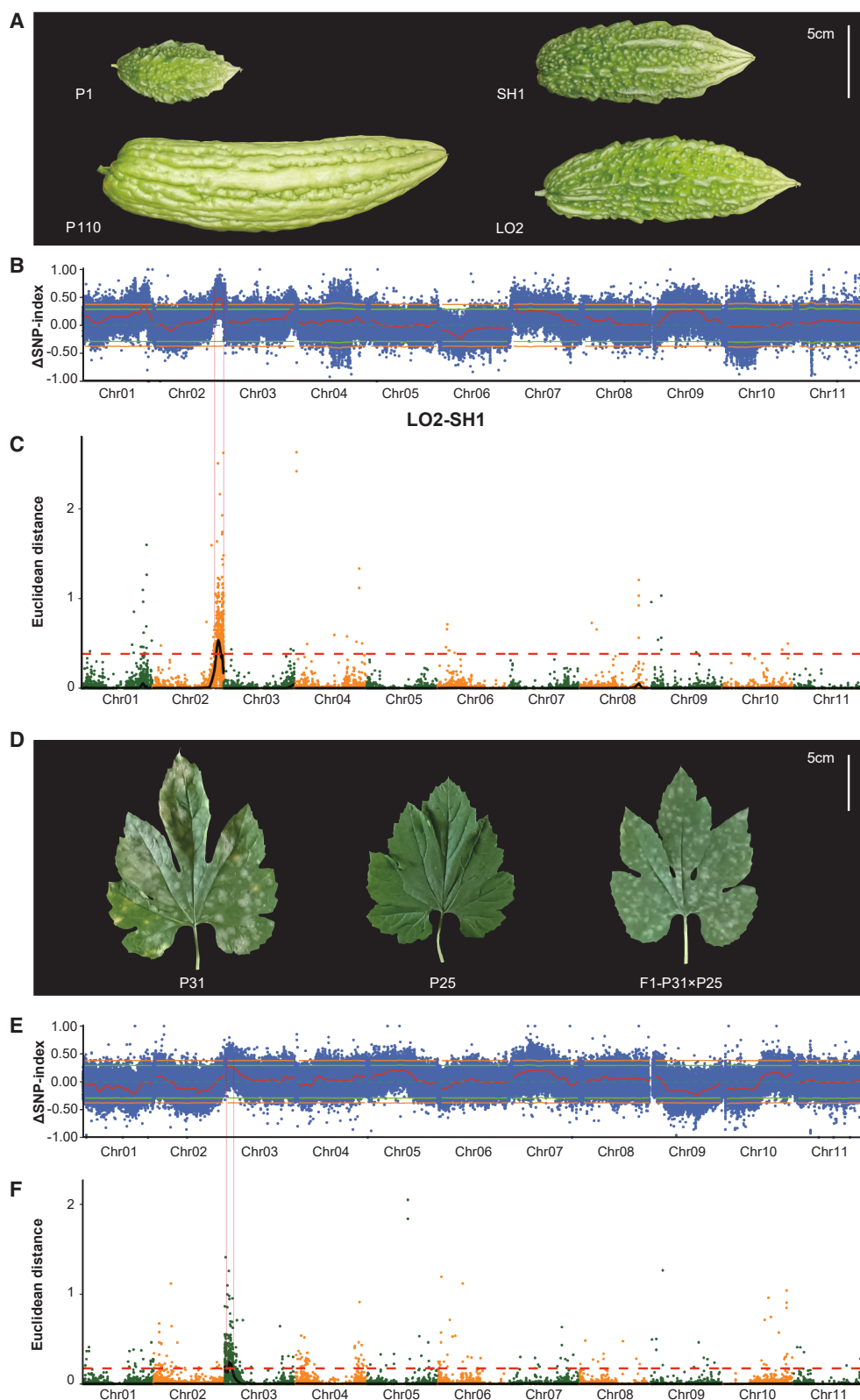


Figure 5. BSA and BSR analyses of bitter melon fruit length and powdery mildew.

(A) Phenotypes of bitter melon fruits from four BSA pools. Scale bar, 5 cm.

(B) Distribution of SNP-index association values from BSA pooled samples across chromosomes. The horizontal axis indicates chromosome names. Colored dots denote calculated Δ SNP-index values, and the red line represents the smoothed Δ SNP-index curve. The green line shows the 95% threshold, and the orange line indicates the 99% percentile threshold.

(legend continued on next page)

genetic mapping analyses focused on three key traits in bitter gourd.

Fruit length

P1 is a near-wild germplasm with a fruit length of 9–12 cm, whereas P110 is an elite breeding parent with a fruit length of 32–38 cm. To identify key genes that regulate fruit length, we conducted a cross between P1 and P110. The F_1 generation exhibited fruit lengths of approximately 18–25 cm (Figure 5A). Subsequently, we established an F_2 segregating population that consisted of 330 lines. Within this F_2 population, 43 plants displayed fruit lengths of no more than 12 cm, whereas 46 plants exhibited fruit lengths greater than 28 cm. These observations suggest that the long-fruit phenotype shows semi-dominance over the short-fruit phenotype. We selected two groups of samples with extremely contrasting phenotypes from the F_2 mapping population; these groups were subjected to BSA and BSR (Supplemental Table 19). Through this combined approach, we successfully identified genomic regions and candidate loci associated with these important traits. Our fruit-length analysis compared contrasting phenotypes, including short oval fruits (P1 and SH1) and elongated fruits (P110 and LO2), as illustrated in the figure. Δ SNP-index and Euclidean distance (ED) analyses identified a prominent peak on Chr02 within the region from 28.9 to 32.4 Mb that exceeded the significance threshold, strongly suggesting that this region harbors genetic determinants controlling fruit-length variation (Figures 5B and 5C). This interval spans 3.5 Mb and contains 2445 variants (above the 0.99 threshold), as well as 371 genes. Further filtering for variants that cause amino acid substitutions identified 91 candidate variants across 74 genes (Supplemental Table 20). Through transcriptomic differential expression analysis and functional annotation, we detected six candidate genes within this region that potentially regulate the fruit-length trait. Subsequent functional analysis identified Y52_Chr0202842 as a likely candidate gene. During qPCR assays, Y52_Chr0202842 exhibited higher expression levels in P110 than in P1. Furthermore, we selected nine bitter gourd germplasms with fruit lengths exceeding 37 cm and nine germplasms with fruit lengths shorter than 12 cm; the results indicated that Y52_Chr0202842 was more highly expressed in germplasms with longer fruits (Supplemental Figure 13A). Y52_Chr0202842 encodes a protein containing a sterile alpha motif (SAM) domain, which was originally identified as a prominent interaction domain involved in diverse cellular processes, including protein–protein and protein–RNA interactions (Denay et al., 2017). Recent studies have demonstrated that although the SAM domain does not directly bind DNA, it can substantially influence the DNA-binding landscape of transcription factors (Olasz et al., 2024). Moreover, homologous genes are predominantly expressed in seeds, flower buds, and carpels in *Arabidopsis* (Denay et al., 2017), where they

regulate grain length and panicle weight through modulation of downstream target gene expression in rice (Liu et al., 2025a, 2025b). Similarly, in bitter gourd, Y52_Chr0202842 exhibits elevated expression in P110, a genotype characterized by longer fruits. Taken together, these findings suggest that Y52_Chr0202842 is a strong candidate gene associated with fruit-length variation.

Fruit tubercles

For the fruit tubercle trait, we performed a cross between P25, which exhibits abundant tubercles, and P31, which exhibits few tubercles (Supplemental Figure 13D). The F_1 generation displayed numerous tubercles, indicating that the presence of tubercles is dominant over their absence. Subsequently, 329 F_2 individuals were generated; 47 lines with abundant tubercles and 44 lines with few tubercles were selected to construct DNA pools for *de novo* sequencing. We then utilized the same BSA and BSR strategies to investigate genetic determinants underlying fruit tubercle formation (Supplemental Table 21). Contrasting phenotypes were selected from the F_2 mapping population, including smooth fruit (P31) and highly tuberculated fruit (P25) (Supplemental Figure 13D). Δ SNP-index and ED analyses revealed a significant peak on Chr04 spanning the region from 3.1 to 7.4 Mb, exceeding the significance threshold (Supplemental Figures 13E and 13F). This genomic interval of approximately 4.3 Mb contained 4779 variants above the threshold and 626 annotated genes. Among these variants, 275 potentially functional mutations causing amino acid substitutions across 155 genes were identified (Supplemental Table 22). Integration of BSR analysis with transcriptomic differential expression profiling and functional annotation subsequently enabled the identification of 22 candidate genes potentially associated with the fruit tubercle trait. Notably, a 1-bp insertion was detected in the start codon region of Y52_Chr0400763, which caused a frameshift mutation and introduced a premature stop codon (Supplemental Figure 13G). For this locus, 18 tubercle-bearing and 35 non-tubercle-bearing bitter gourd germplasm accessions were selected for analysis. The 1-bp insertion co-segregated with the absence of tubercles and demonstrated a phenotypic prediction accuracy of 98% (Supplemental Table 23). Y52_Chr0400763 encodes an epidermal patterning factor 2-like protein belonging to the EPIDERMAL PATTERNING FACTOR-LIKE (EPFL) family. The functions of homologous genes have been studied in *Oryza sativa* and *Arabidopsis*. In both species, EPFL proteins regulate traits associated with epidermal cell division (i.e., awn development in rice and integument formation in *Arabidopsis*) (Bessho-Uehara et al., 2016; Zhang et al., 2019; Xiong et al., 2022; Li et al., 2023). Given that fruit tubercle formation is also associated with cell division and Y52_Chr0400763 shows high expression in the P25 germplasm across multiple developmental stages, particularly from S1 to S3 (Supplemental

- (C) Distribution of ED association values from BSR analysis across chromosomes. The horizontal axis indicates chromosome names. Colored dots denote the ED value at each SNP site, the black line represents the smoothed ED values, and the red dashed line indicates the significance threshold.
- (D) Phenotypes of bitter gourd leaves from the parental lines and the F_1 generation used in BSA. Scale bar, 5 cm.
- (E) Distribution of SNP-index association values from BSA pooled samples across chromosomes. The horizontal axis indicates chromosome names. Colored dots denote calculated Δ SNP-index values, and the red line represents the smoothed Δ SNP-index curve. The green line shows the 95% threshold, and the orange line indicates the 99% percentile threshold.
- (F) Distribution of ED association values from BSR analysis across chromosomes. The horizontal axis indicates chromosome names. Colored dots denote the ED value at each SNP site, the black line represents the smoothed ED values, and the red dashed line indicates the significance threshold, set at 0.173998.

Plant Communications

Figure 13B), Y52_Ch0400763 is likely the candidate gene responsible for fruit tubercle formation.

Powdery mildew resistance

Powdery mildew, one of the major diseases affecting bitter melon, can lead to substantial yield losses. However, the identification and characterization of resistance genes and associated molecular markers remain limited. From the bitter melon germplasm collection, we conducted a cross between a resistant line (P25) and a susceptible line (P31). The F_1 generation exhibited moderate resistance to powdery mildew. In the F_2 generation, 49 plants showed no powdery mildew infection, 66 plants were highly susceptible, and 214 plants exhibited moderate resistance. These results suggest that resistance to powdery mildew displays semi-dominant inheritance (Figure 5D). To investigate the genetic basis of powdery mildew resistance, we generated bulked pools from resistant and susceptible individuals derived from the F_2 mapping population, then conducted BSA and BSR (Supplemental Table 24). The resistant (P25) and susceptible (P31) parental lines, together with their F_1 population, clearly showed contrasting powdery mildew phenotypes (Figure 5D). Δ SNP-index and ED analyses consistently identified a significant genomic region on Chr03 spanning 1.5 to 3.3 Mb, which exceeded the significance threshold (Figures 5E and 5F). This candidate interval, covering approximately 1.5 Mb, contained 440 variants significantly associated with resistance and 123 annotated genes. After filtering for variants predicted to cause amino acid substitutions, 17 candidate variants across 16 genes remained (Supplemental Table 25). Subsequent transcriptomic differential expression analysis combined with functional annotation identified two candidate genes potentially associated with powdery mildew resistance. Further functional analysis indicated that Y52_Ch0300357 is the most likely candidate gene. This gene belongs to the cytochrome P450 family and exhibits higher expression levels in susceptible plants. In plants, cytochrome P450 monooxygenases (CYPs) participate in numerous metabolic pathways that contribute to the synthesis and modification of diverse metabolites with critical functions throughout plant development, including defense responses, signal transduction, organ patterning, and the biosynthesis of structural polymers (Distéfano et al., 2021). Previous studies in cucumber showed that a P450 family gene, CYP82D47, is associated with powdery mildew resistance (Wang et al., 2024a, 2024b). In *Brachypodium distachyon*, overexpression of *BdCYP711A29* increased susceptibility to *Fusarium graminearum* (Changenet et al., 2021). Consistent with these findings, Y52_Ch0300357 showed significantly higher expression in P31 than in P25 after inoculation with powdery mildew (Supplemental Figure 13C). Sequence alignment revealed 24% identity and 44% similarity between Y52_Ch0300357 and *BdCYP711A29*. Therefore, we propose that Y52_Ch0300357 is a potential candidate gene associated with powdery mildew susceptibility in bitter melon.

Comprehensive transcriptomic and metabolomic landscape of bitter melon

To comprehensively characterize tissue-specific gene expression and metabolite profiles in bitter melon, we performed integrated transcriptomic and metabolomic analyses across four distinct tissues from nine diverse bitter melon cultivars (Figure 6A). t-distributed stochastic neighbor embedding

Bitter melon T2T genome and mutation library

(t-SNE) of the transcriptomic and metabolomic datasets revealed clear clustering of leaf, flower, and fruit tissues into distinct groups. However, samples from young and mature fruits showed considerable overlap, indicating strong similarity between these developmental stages (Supplemental Figures 14A and 14B).

Transcriptome profiling identified a total of 25 189 expressed genes (i.e., genes with transcripts per million (TPM) > 1), which were subsequently categorized according to their expression patterns (Supplemental Figure 14D). On average, 16 726 genes were expressed per tissue per variety (Supplemental Figure 14C). To systematically define expression patterns and evaluate tissue specificity and variety-associated shifts, we classified genes into eight distinct categories. A core set of 8407 genes exhibited stable and robust expression across all tissues and varieties (Figure 6B); Gene Ontology (GO) enrichment analysis indicated that these genes mainly function in fundamental cellular and metabolic processes (Supplemental Figure 15A). Notably, leaf tissues contained the highest numbers of preferentially expressed genes (3804 [15.10%]) and tissue-specific genes (1243 [4.93%]), followed by fruit and flower tissues (Figure 6B). GO enrichment analysis of tissue-specific genes revealed substantial functional enrichment in pathways related to development and hormone signaling (Supplemental Figures 15B–15D). Hierarchical clustering of gene expression profiles clearly separated the four tissue types and revealed distinct expression patterns among core, preferentially expressed, and tissue-specific gene groups (Figure 6C).

Metabolomic analysis identified diverse classes of metabolites, with phenolic acids (17.57%), flavonoids (12.26%), lipids (11.38%), and alkaloids (10.85%) representing the most abundant groups (Figure 6D). We classified metabolites into seven categories based on their accumulation patterns (Supplemental Figure 14E). Among the detected metabolites, 40.03% (679 metabolites) exhibited core accumulation patterns across all tissues (Figure 6E). Tissue-preferred metabolites included leaf-preferred (15.47%, 262 metabolites), flower-preferred (12.69%, 215 metabolites), and fruit-preferred (12.69%, 215 metabolites); tissue-specific metabolites comprised leaf-specific (11.82%, 200 metabolites), fruit-specific (4.66%, 79 metabolites), and flower-specific (2.63%, 45 metabolites) (Figure 6E). These findings highlight distinct transcriptomic and metabolomic signatures associated with specific bitter melon tissues, thus providing valuable insights into the molecular basis of tissue-specific functions and biochemical pathways.

Overall, systematic analysis of gene expression patterns and metabolite profiles across multiple tissues from diverse bitter melon varieties establishes a comprehensive transcriptomic and metabolic landscape for cultivated bitter melon. This integrated multi-omics resource offers important insights and constitutes a foundational framework that may facilitate future studies of gene function and metabolic regulation in bitter melon.

To identify genes associated with metabolite accumulation in bitter melon, we performed weighted gene co-expression

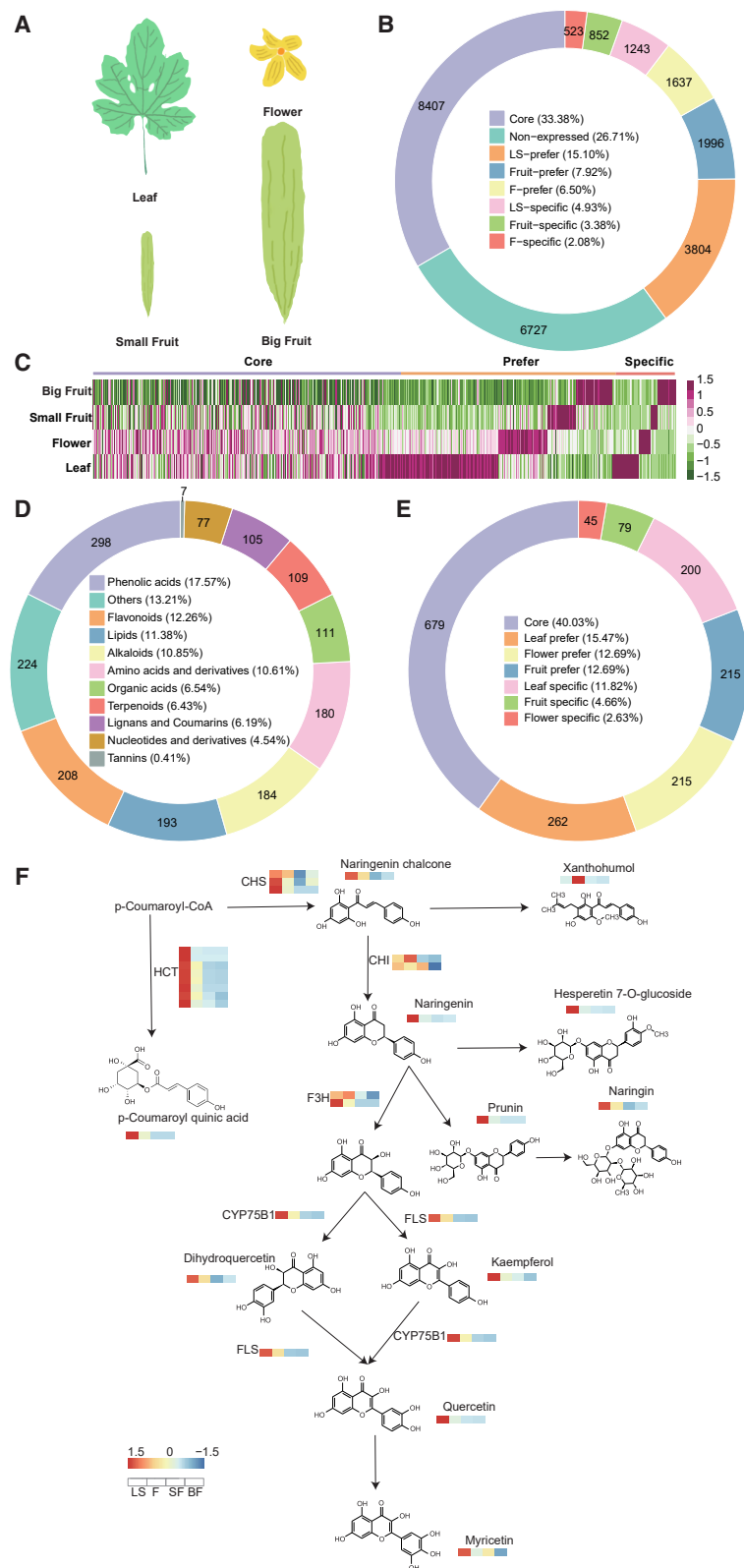


Figure 6. Transcriptomic and metabolomic analyses of different tissues in bitter gourd.

(A) Schematic illustration of the four bitter gourd tissues analyzed: leaf, flower, small fruit, and large fruit.

(B) Distribution of genes according to expression patterns across tissues. Genes are categorized as core (expressed in all tissues), non-expressed, tissue-preferred (leaf, flower, fruit), and tissue-specific (leaf, flower, fruit) groups. Numbers indicate gene counts in each category, with percentages shown in parentheses.

(C) Heatmap showing expression profiles of genes classified as core, tissue-preferred, and tissue-specific across the four tissues. The color scale represents normalized expression levels.

(D) Classification of metabolites identified in bitter gourd tissues. Metabolites are grouped into major biochemical classes, with counts and percentages indicated for each class.

(E) Distribution of metabolites according to accumulation patterns across tissues. Metabolites are classified similarly to genes in **(B)**, including core, tissue-preferred (leaf, flower, fruit), and tissue-specific (leaf, flower, fruit) groups. Numbers indicate metabolite counts, with percentages shown in parentheses.

(F) Heatmaps illustrating differences in the expression levels of genes and metabolites involved in the pathway across four tissues—stems and leaves, flowers, small fruits, and large fruits—arranged from left to right. Gene names are shown on the left side of the heatmaps, and metabolite names appear at the top. HCT, shikimate O-hydroxycinnamoyl transferase; CHS, chalcone synthase; F3H, naringenin 3-dioxygenase; FLS, flavonol synthase; CYP75B1, flavonoid 3'-mono-oxygenase.

enrichment analysis of four highly correlated modules—MEred (BF), MEyellow (SF), MEblack (F), and MEbrown (LS)—indicated their involvement in critical metabolic pathways (Supplemental Figure 16C). The MEred module, enriched in amino acid biosynthesis, contained genes involved in β -alanine and histidine metabolism, which regulate pantothenate synthesis, seed oil accumulation, and stress resistance (Miret and Munné-Bosch, 2014; Ma and Wang, 2016; Seo et al., 2016; Parthasarathy et al., 2019; Sun et al., 2024). The MEyellow module showed enrichment in brassinosteroid biosynthesis; it mediates crosstalk between jasmonic acid and brassinosteroids, thus balancing plant growth and defense response (Wang et al., 2021). The MEblack module, enriched in starch and sucrose metabolism, included genes associated with D-sorbitol biosynthesis, a pathway that regulates osmotic stress tolerance and developmental processes (Jia et al., 2024; Liu et al., 2025a). Notably, the MEbrown module was enriched in phenylalanine metabolism, which represents the biosynthetic origin of phenolic compounds such as salicylic acid, caffeic acid, and quercetin. These compounds participate in plant immunity and exhibit vital pharmacological activities

network analysis (WGCNA) using TPM values, which revealed 13 modules that displayed distinct tissue-specific correlations with key metabolites (Supplemental Figures 16A and 16B). Kyoto Encyclopedia of Genes and Genomes (KEGG)

(Dixon et al., 2002; Demirci et al., 2020; Lefevre et al., 2020; Li et al., 2021a, 2021b, 2022a, 2022b; Oosalo et al., 2024; Vollmannová et al., 2024). Further analysis of the flavonoid biosynthetic pathway revealed elevated expression

of naringenin, kaempferol, and chalcone synthase (CHS) in leaves (Figure 6F). CHS, the rate-limiting enzyme in flavonoid biosynthesis, is induced under biotic and abiotic stresses; it contributes to phytoalexin accumulation and salicylic acid-mediated defense responses (Dao et al., 2011). The high abundance of bioactive flavonoids and expression of their biosynthetic genes in bitter melon leaves and stems—combined with their documented antioxidant, anti-inflammatory, and disease-preventive properties (Motallebi et al., 2022; Shen et al., 2022; Bangar et al., 2023)—establishes bitter melon as a promising source of flavonoid-based nutraceuticals and therapeutic compounds.

Bittergourd DB: An integrated bitter melon resource database

By integrating all datasets generated in this study, we developed a comprehensive bitter melon resource database, Bittergourd DB (<http://bittergourddb.omicsplant.cn/>). The database comprises five main modules: Download, Mutant, Variation, Expression, and Metabolome. In the Download module, users can access the Y52 reference genome and its annotation files. The Mutant module provides phenotype images derived from the mutant library, categorized according to trait variation categories. The Variation module contains variant sites identified from whole-genome resequencing of 320 mutants; users can retrieve all variants within a specified genomic interval by entering the corresponding coordinates. The Expression and Metabolome modules provide gene expression levels and metabolite abundance data for four tissues across nine samples. After selecting a target tissue, users can query a specific gene or metabolite to obtain an abundance table and a corresponding expression heatmap across the nine accessions. Overall, Bittergourd DB provides an integrated and user-friendly platform that facilitates gene discovery, functional validation, and molecular breeding in bitter melon.

DISCUSSION

This study presents a T2T gap-free genome assembly for the elite bitter melon inbred line Y52, which substantially surpasses previous assemblies—including the Mca reference genome—in completeness. By integrating PacBio HiFi long-read sequencing, ONT ultra-long reads, and Hi-C chromosome conformation capture data, we successfully resolved previously inaccessible genomic regions, such as highly repetitive sequences and complex structural elements. This integration resulted in exceptional assembly completeness (genome BUSCO: 98.50%) and base-level accuracy (QV: 55.25), establishing a robust foundation for downstream functional genomic analyses (annotation BUSCO: 98.30%). In addition to serving as a high-quality reference for gene annotation and variant identification, this T2T assembly provides critical insights into genomic architecture that were previously unavailable, including the complete resolution of centromeric regions and accurate representation of telomeric sequences. High-quality genome assemblies are widely regarded as essential prerequisites for precise candidate gene discovery, characterization of structural variation, and development of reliable molecular markers for breeding applications. However, even a high-quality single reference genome cannot capture the full extent of genetic diversity within a species (Shi et al.,

2023). Recently, considerable progress has been achieved in pan-genome studies within the Cucurbitaceae family, greatly advancing the understanding of genetic diversity and structural variation in these species (Li et al., 2022a, 2022b; Lyu et al., 2023; Zhang et al., 2024). Thus, there is an urgent need to integrate diverse germplasm resources worldwide to construct a bitter melon pan-genome, comprehensively characterize its genomic landscape, identify rare allelic variants, and provide a more complete genomic foundation for efficient utilization of germplasm resources and molecular breeding.

A pivotal discovery facilitated by this T2T assembly is the comprehensive characterization of bitter melon centromeres. Through ChIP-seq analysis targeting the McCENH3 protein, we precisely defined the centromeric boundaries of all 11 chromosomes. Notably, bitter melon centromeres lack detectable satellite repeat sequences and are instead predominantly composed of LTR retrotransposons from the Gypsy and Copia superfamilies. This structural organization substantially differs from observations in cucumber (centromeres comprise large arrays of 177-bp satellite repeats interspersed with LTR elements), but closely resembles the arrangement in watermelon, which also lacks centromeric satellite repeats and is enriched in Gypsy-like retrotransposons (Zhang et al., 2024). A third structural pattern occurs in melon and pumpkin, where some centromeres are dominated by TEs, whereas others are dominated by centromeric satellite sequences (Figure 2 and Supplemental Figure 3). Such extensive structural divergence within a single family is remarkable, particularly given that these species diverged only approximately 50–60 million years ago (Guo et al., 2020). The transition from MITE elements to satellite repeats observed in cucumber implies a mechanism for centromere evolution. Our results suggest that TEs can serve as seed sequences that subsequently evolve into satellite arrays through amplification and homogenization, a process similar to the 45S rDNA-driven centromere remodeling reported in tobacco (Chen et al., 2024).

The divergence of centromere structure among closely related Cucurbitaceae species indicates lineage-specific evolutionary trajectories, potentially driven by differential amplification or removal of repetitive elements, distinct recombination landscapes, and lineage-specific selective pressures. Such evolutionary flexibility may enable cucurbit species to adapt to different genomic environments while maintaining essential centromeric functions. This diversity highlights the centromere paradox: rapid evolution of centromeric DNA despite conservation of centromere function (Henikoff et al., 2001). Notably, compared with similar centromere diversity reported in Solanaceae (Chen et al., 2024), the present findings may represent an example of convergent evolution in angiosperms, implying that centromere structural diversification is a common feature of plant genome evolution.

Whether the satellite-free, retrotransposon-rich centromere structure of bitter melon confers specific advantages or limitations for genome stability, chromosome segregation, or environmental adaptation remains an open question warranting further investigation. In particular, whether TE-dominated centromeres are more susceptible to mis-segregation during meiosis compared with satellite-dominated centromeres requires validation through cytological experiments that directly assess

chromosome behavior. We acknowledge that current centromere annotation mainly relies on ChIP-seq and repeat sequence analysis. Although CENH3 ChIP-seq provides reliable functional evidence for centromere localization, future studies incorporating fluorescence *in situ hybridization* or immunofluorescence microscopy would strengthen these conclusions by allowing direct visualization of centromeric chromatin organization. Comparative analyses that include additional cucurbit species, together with population-level evaluation of centromere variation, will be essential to elucidate the evolutionary forces that shape centromere structure and clarify their functional implications for chromosome biology.

The large-scale EMS-induced mutant library established in this study represents another valuable resource for bitter gourd functional genomics. Through resequencing of 320 mutant lines, we identified approximately 481 184 induced mutations that affect more than 72% of annotated coding genes, substantially expanding the genetic variation available for functional studies. The successful identification of *McEGY1* as the causal gene responsible for leaf and fruit yellowing through MutMap analysis clearly demonstrates the utility of integrating high-quality reference genomes with comprehensive mutant resources. In addition to protein-coding genes, recent studies have shown that a considerable proportion of non-coding sequences in cucurbit genomes are conserved and contain *cis*-regulatory elements associated with agronomic traits (Xin et al., 2024). These regulatory regions—including promoters, enhancers, and other *cis*-acting elements—can strongly influence gene expression patterns linked to agronomic performance. Although EMS primarily induces random point mutations, a subset of these mutations inevitably occurs within regulatory regions, potentially disrupting transcription factor binding sites or altering promoter activity strength. In the future, genome-wide chromatin accessibility can be examined through integration of ATAC-seq and DNase-seq data, combined with ChIP-seq analysis of histone modifications such as H3K4me3 and H3K27ac to accurately identify active promoter and enhancer regions. The combination of the mutant library with systematic annotation of *cis*-regulatory elements offers strong potential for identifying regulatory variants that control quantitative traits and for discovering novel crop improvement targets beyond coding sequences. This strategy will complement conventional CDS-centered analyses and support a more comprehensive understanding of the genetic architecture underlying complex phenotypes in bitter gourd.

Although conventional quantitative trait locus (QTL) mapping can define chromosomal intervals, these regions often encompass hundreds of genes, which hinders direct application in precision breeding. Therefore, identification of candidate genes through functional annotation of homologous genes and integration of high-throughput omics datasets represents an important strategy for elucidating genetic regulatory networks and facilitating targeted trait improvement. By combining genetic mapping with candidate gene analysis, we identified putative regulatory genes associated with each trait, thus providing targets for molecular breeding. Moreover, the present results clarify broader principles underlying the genetic regulation of fruit development and disease responses in cucurbits. The polygenic control of fruit length and the identification of Y52_Chr0202842, a homolog of the rice grain-length regulator RFL, suggest conservation of mechanisms

governing organ-size regulation across distantly related plant families. This finding not only provides a candidate gene for bitter gourd improvement but also establishes a comparative framework for investigating whether SAM-domain-containing proteins have a conserved role in regulating cell division or elongation in fleshy fruits and cereal grains. Concerning tubercles, disruption of an EPFL family gene (Y52_Chr0400763/MC04G1399) in tubercle-less variants underscores the genetic regulation of epidermal patterning. Members of the EPFL family encode secreted peptides that coordinate epidermal cell development; this mutation represents a clear example of how a single genetic alteration can eliminate a complex morphological trait with important post-harvest implications. Regarding powdery mildew resistance, the preliminary association of a cytochrome P450 gene (Y52_Chr0300357) with susceptibility, rather than resistance, is particularly noteworthy. Elevated expression of this gene in susceptible lines is consistent with observations in *Brachypodium*, suggesting that certain P450 enzymes can be exploited by pathogens or participate in metabolic pathways that inadvertently weaken plant defense mechanisms. This finding implicates the gene as a potential susceptibility factor, and its downregulation may represent a novel strategy for improving disease resistance. Overall, this study links economically important traits to specific genomic loci and proposes compelling candidate genes with biologically plausible functions. Experimental validation of these candidates will confirm their functional roles and provide the essential foundation for precision breeding, ultimately enabling development of bitter gourd varieties with optimized fruit size, enhanced post-harvest quality, and durable disease resistance. Given the absence of an efficient genetic transformation system in bitter gourd, the functions of genes associated with fruit length, tubercle formation, and powdery mildew resistance remain unverified. Future studies will focus on functional characterization of these candidate genes to facilitate their application in molecular breeding programs.

Integration of BSA and BSR with the T2T reference genome enabled precise delineation of genomic intervals associated with fruit length, fruit tubercles, and powdery mildew resistance. Identification of candidate genes and development of linked molecular markers provide valuable resources for marker-assisted selection in breeding programs. These results illustrate how high-quality reference genomes can accelerate translation of genetic discoveries into practical breeding applications and facilitate efficient pyramiding of favorable alleles for multiple traits. However, a stable and efficient genetic transformation system has not yet been established for bitter gourd, which partly limits functional validation of candidate genes. Therefore, the development of an efficient genetic transformation and gene-editing system for bitter gourd should become a priority in future research—such advances would greatly advance functional genomics studies and molecular design breeding in bitter gourd. Looking ahead, expansion of genetic mapping populations and integration of diverse germplasm resources will help refine QTL intervals and may reveal additional loci involved in trait variation. Furthermore, integration of multi-omics datasets—including transcriptomics, proteomics, and metabolomics—with the genomic resources generated in this study will provide deeper mechanistic insights into trait regulation. Comparative analyses with related cucurbit species may also uncover conserved genetic modules that control shared

agronomic traits, thereby enabling cross-crop translational applications.

In summary, this study establishes a comprehensive genomic and functional resource platform for bitter melon that encompasses a fully resolved T2T genome assembly, detailed centromere characterization, extensive mutant resources, and candidate genes associated with key agronomic traits. Despite certain limitations, such as the absence of cytological validation for centromere annotation and the need for functional verification of candidate genes, the resources and insights generated here provide a strong foundation for future research. Integration of these genomic resources into breeding programs will accelerate genetic improvement; the methodological framework developed in this study offers a valuable reference for genomic research and targeted breeding strategies in other economically important cucurbit crops.

METHODS

Plant material preparation and sequencing

Extracted genomic DNA was subjected to dual-platform sequencing using PacBio Sequel systems. For PacBio sequencing, DNA samples were processed at BerryGenomics to construct circular consensus sequencing libraries, with an optimal insert size of approximately 15 kb. Next, these libraries were sequenced on the PacBio Sequel platform to generate HiFi reads. The resulting HiFi reads, characterized by high accuracy, served as the primary dataset for subsequent genome assembly.

For Hi-C library preparation and sequencing, high-quality genomic DNA was isolated from juvenile leaf tissues. Chromatin structure was cross-linked and stabilized via formaldehyde, followed by *in situ* Hi-C chromosome conformation capture in accordance with established protocols (Belton et al., 2012). The resultant Hi-C libraries were sequenced on an Illumina NovaSeq 6000 platform (<https://www.illumina.com.cn/systems/sequencing-platforms/novaseq.html>), utilizing a 150-bp paired-end sequencing strategy to ensure high-quality data acquisition.

Ultra-long read sequencing libraries for ONT were prepared via the SQK-LSK110 ligation sequencing kit using the standard protocol (<https://nanoporetech.com/document/genomic-dna-by-ligation-sqk-lsk110>). The purified DNA library was then loaded onto a pre-primed R9.4.1 Spot-On Flow Cell and sequenced using a PromethION platform (<https://nanoporetech.com/products/sequence/promethion>), a high-throughput sequencing system designed for large-scale genomic projects. The ultra-long reads generated through this ONT sequencing strategy were particularly valuable for *de novo* genome assembly and enhanced assembly continuity.

To perform genome annotation, we collected various tissues from Y52 (leaves and stems, flowers, small fruits, and large fruits) for RNA sequencing (RNA-seq). Three biological replicates were prepared for each tissue.

ONT adaptive sampling and ultra-long sequencing to fill the Chr06 telomere

Ultra-long DNA extraction and quality assessment were performed using Nanodrop, Qubit, and pulsed-field gel electrophoresis. Libraries were constructed using the optimized SQK-ULK114 kit protocol. To resolve the unknown sequence at the right terminus of Chr06, this study established a dual-platform integrated strategy that combined nanopore adaptive sampling (AS) with ultra-long read sequencing. For the AS platform, we designed four buffer regions (Chr06: 25 235 964–25 335 964, 25 185 964–25 335 964, 25 035 964–25 335 964, and 24 935 964–

25 035 964) to sequentially enrich the target regions. Each buffer region was paired with a custom FASTA reference sequence that included the target segment and terminal repetitive sequences from other chromosomes to enhance telomere-specific enrichment efficiency. Simultaneously, ultra-long read sequencing used the SQK-ULK114 library preparation kit to generate ultra-long DNA fragments (Lu et al., 2024).

Genome survey

To conduct a comprehensive genome survey, high-throughput sequencing was performed utilizing the Illumina platform (<https://www.illumina.com.cn/>). The resulting short reads were processed for *k*-mer frequency analysis via Jellyfish (v2.3.0) with a specified *k*-mer length parameter (*-m*) of 19 nt (Marçais and Kingsford, 2011). The generated *k*-mer spectra were subsequently analyzed with GenomeScope1.0 to comprehensively identify key genomic features (Vurture et al., 2017), including genome size and heterozygosity rate.

Genome *de novo* assembly

Genome assembly of Y52 was conducted through a four-step approach. (1) NextDenovo was used for preliminary genome assembly with ultra-long reads (Hu et al., 2024) and the following parameters: genome_size = 380 m, read_type = ont, rerun = 3, and read_cutoff = 50 000. (2) When both HiFi and ultra-long nanopore reads were available, hifiasm (v0.18.4) was utilized with the *-ul* parameter to integrate both data types for a more comprehensive assembly (Cheng et al., 2024). (3) The La Jolla Assembler was implemented with default parameters for genome assembly from HiFi reads (Bankevich et al., 2022). (4) hifiasm (v0.16.1) was applied to assemble genomes from HiFi reads, using the parameters *-write-ec -write-paf -l0* (Cheng et al., 2021). The resulting draft contigs were subsequently assembled into a gap-free genome via proximity-guided assembly (Li et al., 2021a, 2021b). ONT sequencing data were used to fill genomic gaps and restore missing telomeric regions generated during hifiasm assembly.

HiC-Pro was utilized to classify Hi-C data into valid and invalid interaction pairs; only valid pairs were retained for downstream analysis (Servant et al., 2015). EndHiC was then used to assemble large contigs into chromosome-scale scaffolds by leveraging Hi-C interactions derived from contig end regions, rather than entire contigs (Wang et al., 2022). Finally, a genomic interaction heatmap was generated via HiCPlotter to visualize chromosomal organization and interaction patterns (Akdemir and Chin, 2015).

Evaluation of genome assemblies

Genome assembly quality was comprehensively evaluated through multiple metrics. Assembly completeness was assessed via BUSCO analysis against the embryophyta_odb10 database (Manni et al., 2021). Genome continuity was evaluated by calculating contig N50 length and examining mapping coverage using Qualimap2 (Okonechnikov et al., 2016). Assembly accuracy was quantified through sequencing read mapping rates and consensus QV calculations (Rhie et al., 2020). Additionally, the assembly level was evaluated using the LAI (Ou et al., 2018), which leverages repeat sequence analysis to further validate genome assembly quality.

Genome annotation

Genome repetitive elements were annotated using the Extensive *de novo* TE Annotator (EDTA) (Ou et al., 2019). We first conducted *de novo* identification of TEs with the parameters *-species others -sensitive 1 -anno 1 -evaluate 1*, which generated specialized TE libraries for subsequent structural annotation. To improve annotation accuracy, we implemented RepeatMasker with the parameters *-e rmbblast -s -gff -nolow -no_is -norna* to mask the genome and annotate TE elements using a high-quality, non-redundant TE library (Tarailo-Graovac and Chen, 2009).

Structural annotation of protein-coding genes was performed using the gene prediction softwares AUGUSTUS and GeneMark-ETP, integrated within the BRAKER3 pipeline (Stanke et al., 2008; Brůna et al., 2024; Gabriel et al., 2024). The BRAKER3 workflow incorporated predictions derived from three sources: *ab initio* gene prediction, transcriptional evidence from RNA-seq data, and homologous protein information. A soft-masked version of the genome assembly was generated using the EDTA (Ou et al., 2019) make_masked.pl script with default parameters. RNA-seq reads were aligned to the Y52 genome using HISAT2 (Kim et al., 2019) with default settings, producing sequence alignment/map (SAM) files that were subsequently converted to binary alignment/map (BAM) format and then sorted via SAMtools (Danecek et al., 2021). Protein sequences from the Viridiplantae odb12 protein dataset (https://bioinf.uni-greifswald.de/bioinf/partitioned_odb12/) were provided as the protein database for BRAKER3. Functional annotation of protein-coding genes was accomplished through four complementary approaches: (1) BlastP (Altschul et al., 1997) searches against the NCBI non-redundant protein database (Finn et al., 2008) and Swiss-Prot database (Bairoch and Apweiler, 1999); (2) identification of protein domains and GO terms using InterProScan (Jones et al., 2014); (3) pathway annotation via the KEGG Automatic Annotation Server (Moriya et al., 2007); and (4) orthology-based functional annotation using eggNOG-mapper (Cantalapiedra et al., 2021).

To obtain more comprehensive genomic information, we also annotated non-coding RNAs. Ribosomal RNA genes were identified using the Euk library of Barrnap. snoRNAs were annotated via ShortStack with default parameters (Axtell, 2013). tRNAs were annotated using the eukaryotic library (-E) of tRNAscan-SE (Lowe and Eddy, 1997).

ChIP-seq and analysis

We screened the McCENH3 gene Y52_Chr1101067 and synthesized the corresponding antibody at PhytoAB. ChIP assays were then performed using the generated anti-CENH3 antibodies, as previously described (Tian et al., 2025). ChIP-enriched and input DNA samples were used for library construction according to protocols provided by the VAHTS Universal DNA Library Prep Kit for Illumina V4 (Vazyme, ND610) and the VAHTS Multiplex Oligos Set 4 for Illumina (Vazyme, N321). The resulting libraries were sequenced on the Illumina NovaSeq 6000 platform using 150-bp paired-end reads. Clean McCENH3 ChIP-seq reads were aligned to the Y52 genome assembly using Bowtie2 (v2.5.4) with default parameters (Langmead and Salzberg, 2012). Alignments with mapping quality scores below 30 were discarded, and duplicate reads were removed using SAMtools (v1.21). CENH3-binding domains were identified by comparing ChIP and input samples via MACS2 (v2.2.9.1) (Zhang et al., 2008). We also utilized an integrative approach that combined peaks obtained by mapping Y52 CENH3 ChIP-seq reads to the Mca genome, TE distribution patterns, and synteny with the Y52 genome.

Centromere identification and analysis

Centromeric regions in the Y52 genome were identified using quarTeT (Lin et al., 2023) in combination with CENH3 ChIP-seq data, and satellite sequences were detected using TRASH (Włodzimierz et al., 2023). Results generated by EDTA were used to quantify TE content within centromeric regions. Color-coded identity heatmaps of genomic regions were generated via StainedGlass (Vollger et al., 2022). To address sequence variation among satellite monomers of the same type, the five most representative satellite monomer sequences from each chromosome for each species were selected for phylogenetic analysis. These sequences were chosen based on monomer length and the number of repeat copies present within centromeric intervals.

Construction of mutant library

The inbred line Y52 served as the EMS experimental material in this study. Seeds were provided by the Institute of Tropical Crop Genetic Resources in Hainan Province, China. A total of 10 000 seeds were soaked in ddH₂O

overnight and then treated with 0.2% EMS at room temperature under gentle shaking for 10 h. After treatment, the seeds underwent detoxification with 0.1 mol/l sodium thiosulfate; this step was repeated five times. Subsequently, seeds were washed 15 times with ddH₂O, and each wash lasted 5 min. The treated seeds were wrapped in wet cloth and transferred to an incubator maintained at 37°C under dark conditions until germination occurred. Approximately 3–5 days later, germinated seeds were transplanted into small pots and placed in a greenhouse until 4–6 true leaves had developed. The seedlings were then transplanted into open-field soil. Through artificial self-pollination, 3223 M₁ plants produced M₂ mature seeds, from which 10–60 seeds per plant were harvested, labeled, and stored. For phenotypic evaluation, 10–20 randomly selected M₂ seeds from each M₁ plant were sown in soil, and their morphological traits were recorded.

Methodological criteria for analysis of mutant phenotypes

Dwarf mutants: These mutants were classified based on a clear and severe reduction in plant height (typically half or less) compared with Y52 during the vegetative growth stage. **Leaf color:** For albino and yellow leaves, we determined their characteristic coloration by visual inspection. For light-green leaves, which show only slight differences from Y52, we captured photographs using a Sony (α7RII) digital camera. Three areas were selected to obtain RGB values, which were then calculated using the Green Index (Signorelli et al., 2025). **Bud primordia:** In normal plants, bud primordia at the apical meristem develop into elongated stems, branches, or flowers. In the mutant, bud primordia develop into tendrils. **Bent fruit:** A straight line was drawn between the two terminal ends of the fruit. The fruit was classified as bent when the perpendicular distance between this line and the central axis exceeded 5 cm. **Early senescence:** Compared with Y52, mutant leaves exhibit premature yellowing (chlorosis), beginning in the lower foliage and progressing upward, accompanied by the formation of necrotic lesions. **Female flower defect:** In mutant plants, the ovary of the female flower is considerably shortened. **Reduced lateral branches:** In Y52, nearly every axillary bud develops into a lateral branch during the seedling and vegetative growth stages, whereas the mutant displays greatly reduced branching. **Flat stem:** The stem of Y52 is round, whereas the mutant stem exhibits a flattened morphology, accompanied by excessive branching and clustering. **Warts:** Fruit warts in bitter gourd include grain warts, stripe warts, and flat warts, as defined in the Descriptors and Data Standard for Bitter Gourd (Shen Di, 2008). **Truncate base of fruit:** According to the Descriptors and Data Standard for Bitter Gourd (Shen Di, 2008), the fruit base is categorized into three types—concave, truncate, and convex—based on visual observation. **Nearly round blossom end of fruit:** According to the Descriptors and Data Standard for Bitter Gourd (Shen Di, 2008), the blossom-end shape is classified into three categories—acute tiny, obtuse tiny, and nearly round—based on visual assessment. **Gynocious:** The entire plant produces only female flowers and lacks male flowers. **Piebald leaf:** White or yellow spots on the leaves are visible to the naked eye. **Flower defect:** The entire plant produces no female flowers, and male flowers are either few or numerous at each node. **Rich stem villus:** The stem is densely covered with villi, particularly on vigorously growing stems. **Rolled leaf:** Leaves are not flat and fully expanded but instead display a curled morphology. **Short fruit:** The length of the commercial fruit does not exceed 25 cm. **Spindly fruit:** The blossom end of the fruit exhibits a slender and elongated shape. **Tendrils development:** Tendrils develop into female flowers. **Twin fruit:** A single carpodium produces two fruits. **White/yellow fruit:** The fruit peel appears white or yellow upon visual observation.

Assessment of resistance to powdery mildew

Powdery mildew is a fungal plant disease caused by species within the Erysiphaceae family. It is characterized by the appearance of a white, powdery mycelial growth on the surfaces of leaves, stems, flower buds, and other aerial plant parts. Disease severity was evaluated based on

the percentage of leaf area affected, determined through photographic documentation and visual inspection. The disease grading scale was defined as follows: grade 0, no symptoms observed; grade 1, less than one-third of the leaf area affected, with faint powdery deposits; grade 2, one-third to two-thirds of the leaf area affected, with clearly visible mycelium; grade 3, more than two-thirds of the leaf area affected, with dense powdery coverage and initial chlorosis; and grade 4, extensive mycelial growth accompanied by leaf desiccation or curling. Resistance and susceptibility levels were interpreted as follows: grade 0 indicates resistance; grade 1, slight susceptibility; grade 2, moderate susceptibility; and grades 3 and 4, high susceptibility.

The powdery mildew strain used in this study was *Podosphaera xanthii* race 2 F. The inoculation procedure was performed as follows. Conidia of purified *P. xanthii* race 2 F were obtained from bitter melon strain 31. Leaves exhibiting severe powdery mildew symptoms were collected, and fresh conidia were gently brushed into a 0.05% (v/v) Tween-20 solution. The suspension was thoroughly mixed to ensure homogeneity and generate a uniform spore suspension. Conidial concentration was determined microscopically with a hemocytometer and adjusted to a final concentration of 1.0×10^6 spores/ml. The inoculum was prepared immediately before use. At the developmental stage of six true leaves plus one apical meristem (six leaves and one heart), plants were inoculated in the evening by uniformly spraying the spore suspension onto all six true leaves using a hand-held sprayer. After inoculation, plants were maintained under high-humidity conditions for 48 h, with relative humidity controlled at 86%–91% and temperature maintained between 21°C and 26°C.

Whole-genome resequencing of the 320 mutants

In total, 322 mutant lines were initially selected to identify nucleotide mutations induced by EMS treatment. Two lines were excluded from subsequent analyses because they showed abnormally high levels of variation. Most selected lines consisted of individual plants displaying phenotypic alterations that were readily observable without magnification. Additionally, 80 individuals without visible phenotypic mutations were included to compare mutation rates between lines with and without visible phenotypes. From the mutant library, 27 M₂ and 293 M₃ mutant lines were selected. Each mutant line was sampled individually, and genomic DNA was extracted independently for whole-genome resequencing. Leaf tissues were collected during the flowering and fruiting stages, and genomic DNA was extracted for resequencing. DNA fragments were prepared for library construction using protocols provided by BerryGenomics. Paired-end sequencing was performed on the Illumina HiSeq 2500 platform, generating approximately 30× raw read coverage. Quality control was conducted via Fastp (v0.20.0) to remove low-quality reads containing more than 10% N bases or more than 50% bases with quality scores below Q20 (Chen et al., 2018). The resulting high-quality reads were used for mutant library analysis.

Read mapping and variant detection in the mutant library

Clean reads were aligned to the Y52 genome using the bwa-mem tool with default parameters (Li and Durbin, 2010). The SAM files generated during alignment were converted into BAM files using SAMtools (Danecek et al., 2021), after which the BAM files were sorted and filtered. PCR duplicates were removed using the GATK MarkDuplicates function. Variant sites were detected by generating genomic variant call format (GVCF) files for each sample with the GATK HaplotypeCaller function. All GVCF files were then merged using the GATK Combine GVCFs function to generate a VCF file containing all variant sites from the 322 mutants (McKenna et al., 2010).

Mutation filtration and functional annotation

Raw VCF files were filtered using VCFtools with the parameters –min-al-alleles 2 –max-al-alleles 2 –max-missing 0.8 –minDP 5 –minGQ 5 –mac 50 –minQ 30. Variants detected in the Y52 material were subsequently

removed from the mutant VCF files. Because variants present in the three Y52 samples may represent natural polymorphisms, VCFtools was utilized to eliminate all loci detected in Y52 (Danecek et al., 2011). To further remove residual inherent variation, BCFtools was used to extract genotypes of the 320 mutants at all loci. The number of occurrences of the 0/1 or 1/1 genotype among the 320 mutants was counted; loci observed only once were excluded. The filtered VCF files were used for subsequent analyses. SNPs and InDels were extracted with VCFtools for downstream analysis. To functionally annotate genetic variants in the VCF files, we used SnpEff to construct a SnpEff database based on the Y52 reference genome and its corresponding gene annotation file (Cingolani et al., 2012), ensuring that the annotation information corresponded to the target genome version.

Mutation validation

To validate mutation sites, primers were designed to amplify genomic regions immediately upstream and downstream of each mutation site. The resulting PCR products, approximately 300 bp in length, were sequenced via the Sanger method (Li et al., 2018; Crossley et al., 2020). Sequence data were analyzed with DNA Baser to confirm the presence of mutation sites.

MutMap population construction and analysis

For MutMap analysis aimed at identifying candidate genes associated with the chlorotic phenotype, 30 F₂ chlorotic mutant seedlings were selected, and genomic DNA was extracted from each individual. Equal amounts of DNA from these samples were pooled, and the pooled DNA was subjected to whole-genome sequencing, achieving ~122× sequencing coverage. Additionally, both the WT and mutant parental lines were sequenced with ~60× coverage. Processing of the raw sequencing data followed procedures described in the section “mutation filtration and functional annotation.” Common variant sites present in both parental pools were removed from the generated VCF files; the remaining variants were analyzed using the SNP-index algorithm implemented in MutMap software (Abe et al., 2012). Annotation of SNPs and InDels was performed using SnpEff software. Because the yellow leaf spots of M2750 and the yellow coloration of the bitter melon peel represent recessive mutations, we used the following criteria to select candidate mutation sites: an SNP-index equal to 1, mutation type consistent with typical EMS mutagenesis, and homozygosity of the mutation site in the F₂ mutant phenotype pool. Candidate variant sites were annotated using SnpEff.

Phylogenetic analysis of EGY1

We collected protein sequences from 33 representative plant species spanning major evolutionary lineages—from glaucophytes to angiosperms—along with sequences from nine published Cucurbitaceae species. Through BlastP searches, we identified 33 homologous EGY1 protein sequences across these species. Sequences were aligned via MAFFT (v7.525) (Nakamura et al., 2018), and phylogenetic trees were reconstructed using IQ-TREE (ultra-fast bootstrap approximation with 1000 replicates) (Minh et al., 2020).

BSA and BSR population construction and analysis

Two groups of samples (>30 individuals per group) exhibiting strongly contrasting phenotypes were selected from the F₂ mapping population. Genomic DNA and total RNA were extracted individually from each sample. For BSA, equal amounts of DNA from each group were pooled; similarly, equal quantities of RNA were pooled for BSR analysis. Subsequently, the pooled DNA and RNA samples were subjected to high-throughput sequencing. Processing of the raw sequencing data followed procedures described in the section “mutation filtration and functional annotation.” Raw VCF files were filtered using VCFtools with the parameters –max-missing 1 –min-alleles 2 –max-alleles 2 –minQ 30 –minDP 10 –maxDP 100. Variant sites shared by the two parental lines were removed; SNP-index values for each pooled sample were calculated using QTL-seq (Takagi et al., 2013), with the

parameters `–min-depth 4 –max-depth 100 –min-SNPindex 0.3`. Genes corresponding to variant sites within peak regions were annotated using SnpEff.

RNA-seq was also performed on the F_2 pooled samples from the fruit tubercle, fruit length, and powdery mildew populations. After quality control of the raw data using Fastp, the index of the Y52 genome was constructed with STAR (Dobin et al., 2013), and the Y52 genome served as the reference for read alignment. Sample names were added to the generated SAM files using SAMtools; reads with mapping quality values below 30 were removed using the command `view -q 30 -b` before BAM files were generated. The BAM files were processed with GATK MarkDuplicates and GATK SplitNCigarReads. The GATK HaplotypeCaller function was used to generate GVCF files for each sample. The GATK Combine GVCFs function was then used to merge all GVCF files, yielding a VCF file that contained all variant sites in the BSR population. Filtering conditions for the VCF files were consistent with those used for BSA. ED values of the F_2 pools were calculated, and loess fitting was performed using the MMAPP2 package (Hill et al., 2013), followed by visualization with ggplot2.

Transcriptomic and metabolomic analysis

To systematically characterize tissue-specific gene expression and metabolite profiles in bitter gourd, we performed integrated transcriptomic and metabolomic analyses across four distinct tissues from nine diverse bitter gourd cultivars. Each tissue type was sampled in triplicate to provide independent biological replicates for transcriptomic and metabolomic sequencing.

Expression levels obtained from transcriptomic and metabolomic datasets were averaged. Subsequently, the Rtsne package (Laurens van der Maaten, 2014) was used for t-SNE analysis and visualization. Gene expression levels were classified according to fold-change differences between tissues. Genes with TPM values below 1 were regarded as non-expressed genes. Genes with fold changes greater than 5, when comparing the highest expression levels with average expression levels in other tissues, were classified as tissue-specific genes. Genes with fold changes between 2 and 5 were classified as tissue-preferentially expressed genes. Genes with fold changes of 2 or less were classified as core-expressed genes. Metabolites were classified using the same criteria. Visualization of the gene expression matrix was performed using the pheatmap package (v1.0.12) in R. GO enrichment analysis was performed via the clusterProfiler package (v4.12.6) in R (Wu et al., 2021).

The TAU method was used to evaluate biased expression patterns of TE-inserted and non-TE-inserted genes across different tissues (Yanai et al., 2005). Weighted gene co-expression network analysis (WGCNA) was conducted using the WGCNA package (v1.72) in R (Langfelder and Horvath, 2008) based on TPM values; genes displaying average expression levels below 1 were excluded from analysis.

Statistical analysis

Statistical analyses were performed in R software (v4.3.1). Student's *t*-test was utilized to compare gene expression levels between powdery mildew-resistant and -susceptible bitter gourd plants at different disease stages. Student's *t*-test was also used to compare gene expression levels between long-fruit and short-fruit bitter gourd plants.

DATA AND CODE AVAILABILITY

Raw sequencing data and genome assemblies have been deposited in the National Genomics Data Center (NGDC, <https://ngdc.cncb.ac.cn/>) under BioProject PRJCA036725. The genome assembly and annotation files are also available on Figshare (<https://figshare.com/s/9a76934a5d8d38c742b4>). Datasets generated in this study are accessible through Bittergourd DB (<http://bittergourddb.omicsplant.cn/>). All materials produced in this study are available upon request by e-mail (yurenbo612@163.com).

The code associated with this study is publicly accessible via GitHub (<https://github.com/ZQZHZ/MCT2T>).

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (32460760), the Project of the National Key Laboratory for Tropical Crop Breeding (NKLTCBCXTD35, SKLTCBZRJJ202503), the Major Science and Technology Plan of Hainan Province (ZDKJ2021010), and the Hainan Province Science and Technology Special Fund (ZDYF2024XDNY183). This research was supported by Bosheng Li, Dongdong Lu, and the Single-Cell Omics and Single-Molecule Analysis Platform of the Peking University Institute of Advanced Agricultural Sciences.

ACKNOWLEDGMENTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

R.Y., Y.Y., and X.W. designed the project; R.Y., Y.Y., Y.Z., and H.H. supervised the project; R.Y. and X.W. conducted the EMS mutation experiment; R.Y., Y.Y., Y.N., X.H., J.L., Z.Z., W.G., and L.L. collected the mutant phenotypes; Y.Z. and Y.W. performed all bioinformatics analyses; Y.Z. performed analyses of genome assembly and annotation; Y.W. performed analysis of the mutant library, centromere characterization, and gene mapping; J.T., X.M., and W.L. performed the ChIP experiments and data analysis; Q.Z. and Y.W. conducted the transcriptomic and metabolomic analyses; Y.Z., R.Y., X.W., and Y.W. wrote the draft; and X.Z. and Y.Z. revised the manuscript. All authors contributed to manuscript preparation and read, commented on, and approved the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: June 5, 2025

Revised: November 26, 2025

Accepted: February 12, 2026

Published: February 16, 2026

REFERENCES

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**:174–178. <https://doi.org/10.1038/nbt.2095>.
- Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**:198. <https://doi.org/10.1186/s13059-015-0767-1>.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Axtell, M.J. (2013). ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* **19**:740–751. <https://doi.org/10.1261/ma.035279.112>.
- Bairoch, A., and Apweiler, R. (1999). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**:49–54. <https://doi.org/10.1093/nar/27.1.49>.
- Bankevich, A., Bzikadze, A.V., Kolmogorov, M., Antipov, D., and Pevzner, P.A. (2022). Multiplex de Bruijn graphs enable genome assembly from long, high-fidelity reads. *Nat. Biotechnol.* **40**:1075–1081. <https://doi.org/10.1038/s41587-022-01220-6>.

- Bangar, S.P., Chaudhary, V., Sharma, N., Bansal, V., Ozogul, F., and Lorenzo, J.M. (2023). Kaempferol: A flavonoid with wider biological activities and its applications. *Crit. Rev. Food Sci. Nutr.* **63**:9580–9604. <https://doi.org/10.1080/10408398.2022.2067121>.
- Behera, T.K., Behera, S., Bharathi, L.K., John, K.J., Simon, P.W., and Staub, J.E. (2010). Botany, horticulture, breeding. In *Horticultural Reviews*, J. Janick, ed., 37, pp. 101–141. <https://doi.org/10.1002/9780470543672.ch2>.
- Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**:268–276. <https://doi.org/10.1016/j.ymeth.2012.05.001>.
- Bessho-Uehara, K., Wang, D.R., Furuta, T., Minami, A., Nagai, K., Gamuyao, R., Asano, K., Angeles-Shim, R.B., Shimizu, Y., Ayano, M., et al. (2016). Loss of function at RAE2, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. *Proc. Natl. Acad. Sci. USA* **113**:8969–8974. <https://doi.org/10.1073/pnas.1604849113>.
- Brúna, T., Lomsadze, A., and Borodovsky, M. (2024). GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res.* **34**:757–768. <https://doi.org/10.1101/gr.278373.123>.
- Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**:5825–5829. <https://doi.org/10.1093/molbev/msab293>.
- Changenet, V., Macadré, C., Boutet-Mercey, S., Magne, K., Januario, M., Dalmais, M., Bendahmane, A., Mouille, G., and Dufresne, M. (2021). Overexpression of a Cytochrome P450 Monooxygenase Involved in Orobanchol Biosynthesis Increases Susceptibility to Fusarium Head Blight. *Front. Plant Sci.* **12**:662025. <https://doi.org/10.3389/fpls.2021.662025>.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. <https://doi.org/10.1038/nmeth.1923>.
- Cheng, H., Asri, M., Lucas, J., Koren, S., and Li, H. (2024). Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat. Methods* **21**:967–970. <https://doi.org/10.1038/s41592-024-02269-8>.
- Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**:170–175. <https://doi.org/10.1038/s41592-020-01056-5>.
- Chen, W., Yan, M., Chen, S., Sun, J., Wang, J., Meng, D., Li, J., Zhang, L., and Guo, L. (2024). The complete genome assembly of *Nicotiana benthamiana* reveals the genetic and epigenetic landscape of centromeres. *Nat. Plants* **10**:1928–1943. <https://doi.org/10.1038/s41477-024-01849-y>.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118. *Fly* **6**:80–92. <https://doi.org/10.4161/fly.19695> iso-2; iso-3. .
- Crossley, B.M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M.L., Clement, T., and Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *J. Vet. Diagn. Invest.* **32**:767–775. <https://doi.org/10.1177/1040638720905833>.
- Cui, J., Luo, S., Niu, Y., Huang, R., Wen, Q., Su, J., Miao, N., He, W., Dong, Z., Cheng, J., et al. (2018). A RAD-based genetic map for anchoring scaffold sequences and identifying QTLs in bitter gourd (*Momordica charantia*). *Front. Plant Sci.* **9**:477. <https://doi.org/10.3389/fpls.2018.00477>.
- Cui, J., Yang, Y., Luo, S., Wang, L., Huang, R., Wen, Q., Han, X., Miao, N., Cheng, J., Liu, Z., et al. (2020). Whole-genome sequencing provides insights into the genetic diversity and domestication of bitter gourd (*Momordica* spp.). *Hortic. Res.* **7**:85. <https://doi.org/10.1038/s41438-020-0305-5>.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008. <https://doi.org/10.1093/gigascience/giab008>.
- Dao, T.T.H., Linthorst, H.J.M., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* **10**:397–412. <https://doi.org/10.1007/s11101-011-9211-7>.
- Denay, G., Vachon, G., Dumas, R., Zubieta, C., and Parcy, F. (2017). Plant SAM-Domain Proteins Start to Reveal Their Roles. *Trends Plant Sci.* **22**:718–725. <https://doi.org/10.1016/j.tplants.2017.06.006>.
- Deng, Y., Liu, S., Zhang, Y., Tan, J., Li, X., Chu, X., Xu, B., Tian, Y., Sun, Y., Li, B., et al. (2022). A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. *Mol. Plant* **15**:1268–1284. <https://doi.org/10.1016/j.molp.2022.06.010>.
- Demirci, T., Çelikkol Akçay, U., and Göktürk Baydar, N. (2020). Effects of 24-epibrassinolide and l-phenylalanine on growth and caffeic acid derivative production in hairy root culture of *Echinacea purpurea* L. Moench. *Acta Physiol. Plant.* **42**:66. <https://doi.org/10.1007/s11738-020-03055-7>.
- Distéfano, A.M., Setzes, N., Cascallares, M., Fiol, D.F., Zabaleta, E., and Pagnussat, G.C. (2021). Roles of cytochromes P450 in plant reproductive development. *Int. J. Dev. Biol.* **65**:187–194. <https://doi.org/10.1387/ijdb.200100gp>.
- Dixon, R.A., Achnine, L., Kota, P., Liu, C.J., Reddy, M.S.S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant Pathol.* **3**:371–390. <https://doi.org/10.1046/j.1364-3703.2002.00131.x>.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L.L., et al. (2008). The Pfam protein families database. *Nucleic Acids Res.* **36**:D281–D288. <https://doi.org/10.1093/nar/gkm960>.
- Fu, A., Zheng, Y., Guo, J., Grierson, D., Zhao, X., Wen, C., Liu, Y., Li, J., Zhang, X., Yu, Y., et al. (2023). Telomere-to-telomere genome assembly of bitter melon (*Momordica charantia* L. var. *abbreviata* Ser.) reveals fruit development, composition and ripening genetic characteristics. *Hortic. Res.* **10**:uhac228. <https://doi.org/10.1093/hr/uhac228>.
- Fu, D., Zhou, H., Grimm, B., and Wang, P. (2025). The BCM1-EGY1 module balances chlorophyll biosynthesis and breakdown to confer chlorophyll homeostasis in land plants. *Mol. Plant* **18**:76–94. <https://doi.org/10.1016/j.molp.2024.11.016>.
- Gangadhara Rao, P., Behera, T.K., Gaikwad, A.B., Munshi, A.D., Jat, G.S., and Boopalakrishnan, G. (2018). Mapping and QTL Analysis of Gynoecy and Earliness in Bitter Gourd (*Momordica charantia* L.)

- Using Genotyping-by-Sequencing (GBS) Technology. *Front. Plant Sci.* **9**:1555. <https://doi.org/10.3389/fpls.2018.01555>.
- Garg, V., Bohra, A., Mascher, M., Spannagl, M., Xu, X., Bevan, M.W., Bennetzen, J.L., and Varshney, R.K. (2024). Unlocking plant genetics with telomere-to-telomere genome assemblies. *Nat. Genet.* **56**:1788–1799. <https://doi.org/10.1038/s41588-024-01830-7>.
- Gabriel, L., Bruna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., and Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res.* **34**:769–777. <https://doi.org/10.1101/gr.278090.123>.
- Guo, J., Xu, W., Hu, Y., Huang, J., Zhao, Y., Zhang, L., Huang, C.-H., and Ma, H. (2020). Phylotranscriptomics in Cucurbitaceae Reveal Multiple Whole-Genome Duplications and Key Morphological and Molecular Innovations. *Mol. Plant* **13**:1117–1133. <https://doi.org/10.1016/j.molp.2020.05.011>.
- Han, X., Zhang, Y., Zhang, Q., Ma, N., Liu, X., Tao, W., Lou, Z., Zhong, C., Deng, X.W., Li, D., et al. (2023). Two haplotype-resolved, gap-free genome assemblies for *Actinidia latifolia* and *Actinidia chinensis* shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. *Mol. Plant* **16**:452–470. <https://doi.org/10.1016/j.molp.2022.12.022>.
- Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* **293**:1098–1102. <https://doi.org/10.1126/science.1062939>.
- Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., et al. (2014). Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. *Plant Cell* **26**:1382–1397. <https://doi.org/10.1105/tpc.113.121590>.
- Hill, J.T., Demarest, B.L., Bisgrove, B.W., Gorski, B., Su, Y.C., and Yost, H.J. (2013). MMAPP: Mutation Mapping Analysis Pipeline for Pooled RNA-seq. *Genome Res.* **23**:687–697. <https://doi.org/10.1101/gr.146936.112>.
- Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A.O., Liang, F., Li, J., Sandoval, J.R., Cooper, D.N., Ye, K., et al. (2024). NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**:107. <https://doi.org/10.1186/s13059-024-03252-4>.
- Jander, G., Baerson, S.R., Hudak, J.A., Gonzalez, K.A., Gruys, K.J., and Last, R.L. (2003). Ethylmethanesulfonate Saturation Mutagenesis in *Arabidopsis* to Determine Frequency of Herbicide Resistance. *Plant Physiol.* **131**:139–146. <https://doi.org/10.1104/pp.102.010397>.
- Jia, X., Xu, S., Wang, F., Jia, Y., Qing, Y., Gao, T., Zhang, Z., Liu, X., Yang, C., Ma, F., et al. (2024). Sorbitol mediates age-dependent changes in apple plant growth strategy through gibberellin signaling. *Hortic. Res.* **11**:uhae192. <https://doi.org/10.1093/hr/uhae192>.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- Karube, H., Suzuki, S., Miyagi, N., Taniai, N., Miyake, M., Tarora, K., Urasaki, N., and Matsumura, H. (2023). Quantitative trait loci for sex ratio in monoecious bitter gourd (*Momordica charantia*). *Sci. Hortic.* **321**:112330. <https://doi.org/10.1016/j.scienta.2023.112330>.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**:907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Laurens van der Maaten. (2014). Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**:3221–3245. <https://doi.org/10.5555/2627435.2697068>.
- Lavale, S.A., Mathew, D., Pradeepkumar, T., Joseph John, K., and Joseph, J. (2023). Mapping the QTL and tagging yield traits in bitter gourd (*Momordica charantia* L.) using microsatellite markers. *Biocatal. Agric. Biotechnol.* **47**:102553. <https://doi.org/10.1016/j.bcab.2022.102553>.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **9**:559. <https://doi.org/10.1186/1471-2105-9-559>.
- Lee-Huang, S., Huang, P.L., Nara, P.L., Chen, H.-C., Kung, H.F., Huang, P., Huang, H.I., and Huang, P.L. (1990). MAP 30: a new inhibitor of HIV-1 infection and replication. *FEBS Lett.* **272**:12–18. [https://doi.org/10.1016/0014-5793\(90\)80438-O](https://doi.org/10.1016/0014-5793(90)80438-O).
- Lefevre, H., Bauters, L., and Gheysen, G. (2020). Salicylic Acid Biosynthesis in Plants. *Front. Plant Sci.* **11**:338. <https://doi.org/10.3389/fpls.2020.00338>.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**:589–595. <https://doi.org/10.1093/bioinformatics/btp698>.
- Li, K., Jiang, W., Hui, Y., Kong, M., Feng, L.Y., Gao, L.Z., Li, P., and Lu, S. (2021a). Gapless *indica* rice genome reveals synergistic contributions of active transposable elements and segmental duplications to rice genome evolution. *Mol. Plant* **14**:1745–1756. <https://doi.org/10.1016/j.molp.2021.06.017>.
- Li, S., Pi, J., Zhu, H., Yang, L., Zhang, X., and Ding, W. (2021b). Caffeic Acid in Tobacco Root Exudate Defends Tobacco Plants From Infection by *Ralstonia solanacearum*. *Front. Plant Sci.* **12**:690586. <https://doi.org/10.3389/fpls.2021.690586>.
- Li, M., Lv, M., Wang, X., Cai, Z., Yao, H., Zhang, D., Li, H., Zhu, M., Du, W., Wang, R., et al. (2023). The EPFL-ERF-SERK signaling controls integument development in *Arabidopsis*. *New Phytol.* **238**:186–201. <https://doi.org/10.1111/nph.18701>.
- Li, Q., Qiao, X., Li, L., Gu, C., Yin, H., Qi, K., Xie, Z., Yang, S., Zhao, Q., Wang, Z., et al. (2024). Haplotype-resolved T2T genome assemblies and pangenome graph of pear reveal diverse patterns of allele-specific expression and the genomic basis of fruit quality traits. *Plant Commun.* **5**:101000. <https://doi.org/10.1016/j.xplc.2024.101000>.
- Li, Z., Jiang, L., Ma, Y., Wei, Z., Hong, H., Liu, Z., Lei, J., Liu, Y., Guan, R., and Guo, Y. (2017). Development and utilization of a new chemically-induced soybean library with a high mutation density. *JIPB* **59**:60–74. <https://doi.org/10.1111/jipb.12505>.
- Li, A., Sun, X., and Liu, L. (2022a). Action of Salicylic Acid on Plant Growth. *Front. Plant Sci.* **13**:878076. <https://doi.org/10.3389/fpls.2022.878076>.
- Li, H., Wang, S., Chai, S., Yang, Z., Zhang, Q., Xin, H., Xu, Y., Lin, S., Chen, X., Yao, Z., et al. (2022b). Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nat. Commun.* **13**:682. <https://doi.org/10.1038/s41467-022-28362-0>.
- Lin, Y., Ye, C., Li, X., Chen, Q., Wu, Y., Zhang, F., Pan, R., Zhang, S., Chen, S., Wang, X., et al. (2023). quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic. Res.* **10**:uhad127. <https://doi.org/10.1093/hr/uhad127>.
- Liu, J., Wang, B., Zhang, S., Liu, M., Liu, W., Yang, P., Liang, C., Song, J., Yang, Y., Wang, R., et al. (2025a). Sorbitol promotes the graft

- healing process in pears. *Hortic. Res.* **12**:uhaf168. <https://doi.org/10.1093/hr/uhaf168>.
- Liu, J., Song, Y., Mei, M., Zhao, X., Wan, S., Xun, Q., Meng, Y., An, J., Li, G., Ding, Y., et al. (2025b). Gene editing unlocks superior mutants from once detrimental RFL for enhanced rice yield traits. *Plant J.* **123**:e70454. <https://doi.org/10.1111/tpj.70454>.
- Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964. <https://doi.org/10.1093/nar/25.5.955>.
- Lu, D., Liu, C., Ji, W., Xia, R., Li, S., Liu, Y., Liu, N., Liu, Y., Deng, X.W., and Li, B. (2024). Nanopore ultra-long sequencing and adaptive sampling spur plant complete telomere-to-telomere genome assembly. *Mol. Plant* **17**:1773–1786. <https://doi.org/10.1016/j.molp.2024.10.008>.
- Lu, X., Liu, J., Ren, W., Yang, Q., Chai, Z., Chen, R., Wang, L., Zhao, J., Lang, Z., Wang, H., et al. (2018). Gene-Indexed Mutations in Maize. *Mol. Plant* **11**:496–504. <https://doi.org/10.1016/j.molp.2017.11.013>.
- Lyu, X., Xia, Y., Wang, C., Zhang, K., Deng, G., Shen, Q., Gao, W., Zhang, M., Liao, N., Ling, J., et al. (2023). Pan-genome analysis sheds light on structural variation-based dissection of agronomic traits in melon crops. *Plant Physiol.* **193**:1330–1348. <https://doi.org/10.1093/plphys/kiad405>.
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **38**:4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770. <https://doi.org/10.1093/bioinformatics/btr011>.
- Matsumura, H., Hsiao, M.C., Lin, Y.P., Toyoda, A., Taniai, N., Tarora, K., Urasaki, N., Anand, S.S., Dhillon, N.P.S., Schafleitner, R., et al. (2020). Long-read bitter melon (*Momordica charantia*) genome and the genomic architecture of nonclassic domestication. *Proc. Natl. Acad. Sci. USA* **117**:14543–14551. <https://doi.org/10.1073/pnas.1921016117>.
- Ma, H., and Wang, S. (2016). Histidine Regulates Seed Oil Deposition through Abscissic Acid Biosynthesis and β -oxidation. *Plant Physiol.* **172**:848–848857. <https://doi.org/10.1104/pp.16.00950>.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Misra, B.B., Langefeld, C., Olivier, M., and Cox, L.A. (2019). Integrated omics: tools, advances and future approaches. *J. Mol. Endocrinol.* **62**:R21–R45. <https://doi.org/10.1530/JME-18-0055>.
- Miret, J.A., and Munné-Bosch, S. (2014). Plant amino acid-derived vitamins: biosynthesis and function. *Amino Acids* **46**:809–824. <https://doi.org/10.1007/s00726-013-1653-3>.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**:W182–W185. <https://doi.org/10.1093/nar/gkm321>.
- Mo, C., Wang, H., Wei, M., Zeng, Q., Zhang, X., Fei, Z., Zhang, Y., and Kong, Q. (2024). Complete genome assembly provides a high-quality skeleton for pan-NLRome construction in melon. *Plant J.* **118**:2249–2268. <https://doi.org/10.1111/tpj.16705>.
- Motallebi, M., Bhia, M., Rajani, H.F., Bhia, I., Tabarraei, H., Mohammadkhani, N., Pereira-Silva, M., Kasaii, M.S., Nouri-Majid, S., Mueller, A.L., et al. (2022). Naringenin: A potential flavonoid phytochemical for cancer therapy. *Life Sci.* **305**:120752. <https://doi.org/10.1016/j.lfs.2022.120752>.
- Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science* **374**:840. <https://doi.org/10.1126/science.abi7489>.
- Nakamura, T., Yamada, K.D., Tomii, K., and Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**:2490–2492. <https://doi.org/10.1093/bioinformatics/bty121>.
- Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**:292–294. <https://doi.org/10.1093/bioinformatics/btv566>.
- Olasz, B., Smithers, L., Evans, G.L., Anandan, A., Murcha, M.W., and Vrielink, A. (2024). Structural analysis of the SAM domain of the *Arabidopsis* mitochondrial tRNA import receptor. *J. Biol. Chem.* **300**:107258. <https://doi.org/10.1016/j.jbc.2024.107258>.
- Oosalo, A.A., Naseri, L., Alirezalu, A., Darvishzadeh, R., and Ebrahimi, S.N. (2024). Exogenous phenylalanine application effects on phytochemicals, antioxidant activity, HPLC profiling, and PAL and CHS genes expression in table grapes (*Vitis vinifera* cv. ‘Qzi Ouzum’). *BMC Plant Biol.* **24**:1216. <https://doi.org/10.1186/s12870-024-05934-4>.
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**:275. <https://doi.org/10.1186/s13059-019-1905-y>.
- Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**:e126. <https://doi.org/10.1093/nar/gky730>.
- Panigrahi, I., Behera, T.K., Munshi, A., Dey, S., Gaikwad, A., and Senapati, M. (2024). SSR markers based QTL mapping and genetic analysis for yield and yield-attributing traits in bitter melon (*Momordica charantia* L.). *S. Afr. J. Bot.* **174**:715–734. <https://doi.org/10.1016/j.sajb.2024.09.049>.
- Parthasarathy, A., Savka, M.A., and Hudson, A.O. (2019). The Synthesis and Role of β -Alanine in Plants. *Front. Plant Sci.* **10**:921. <https://doi.org/10.3389/fpls.2019.00921>.
- Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**:3542–3558.e16. <https://doi.org/10.1016/j.cell.2021.04.046>.
- Rao, P.G., Behera, T.K., Gaikwad, A.B., Munshi, A.D., Srivastava, A., Boopalakrishnan, G., and Vinod. (2021). Genetic analysis and QTL mapping of yield and fruit traits in bitter melon (*Momordica charantia* L.). *Sci. Rep.* **11**:4109. <https://doi.org/10.1038/s41598-021-83548-8>.
- Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**:245. <https://doi.org/10.1186/s13059-020-02134-9>.
- Sarao, N.K. (2021). Quantitative Trait Loci Mapping for Earliness, Fruit, and Seed Related Traits Using High Density Genotyping-by-Sequencing-Based Genetic Map in Bitter Melon (*Momordica*

- charantia* L.). *Front. Plant Sci.* **12**:799932. <https://doi.org/10.3389/fpls.2021.799932>.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**:259. <https://doi.org/10.1186/s13059-015-0831-x>.
- Seo, S., Nakaho, K., Hong, S.W., Takahashi, H., Shigemori, H., and Mitsuhashi, I. (2016). I-Histidine Induces Resistance in Plants to the Bacterial Pathogen *Ralstonia solanacearum* Partially Through the Activation of Ethylene Signaling. *Plant Cell Physiol.* **57**:1932–1942. <https://doi.org/10.1093/pcp/pcw114>.
- Shen Di. (2008). *Description Specifications and Data Specifications for Melon Germplasm Resources* (China Agricultural Press).
- Shen, N., Wang, T., Gan, Q., Liu, S., Wang, L., and Jin, B. (2022). Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* **383**:132531. <https://doi.org/10.1016/j.foodchem.2022.132531>.
- Shi, J., Tian, Z., Lai, J., and Huang, X. (2023). Plant pan-genomics and its applications. *Mol. Plant* **16**:168–186. <https://doi.org/10.1016/j.molp.2022.12.009>.
- Signorelli, S., Casaretto, E., Etchemendy-Gamundi, M., Bentancor, M., and Harvey Millar, A. (2025). The Green Index: A Widely Accessible Method to Quantify the Degree of Greenness of Photosynthetic Organisms. *Plant Cell Environ.* **48**:8027–8043. <https://doi.org/10.1111/pce.70102>.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**:637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
- Sun, X., Li, X., Lu, Y., Wang, S., Zhang, X., Zhang, K., Su, X., Liu, M., Feng, D., Luo, S., et al. (2022). Construction of a high-density mutant population of Chinese cabbage facilitates the genetic dissection of agronomic traits. *Mol. Plant* **15**:913–924. <https://doi.org/10.1016/j.molp.2022.02.006>.
- Sun, Y., Zhou, Y., Long, Q., Xing, J., Guo, P., Liu, Y., Zhang, C., Zhang, Y., Fernie, A.R., Shi, Y., et al. (2024). OsBCAT2, a gene responsible for the degradation of branched-chain amino acids, positively regulates salt tolerance by promoting the synthesis of vitamin B5. *New Phytol.* **241**:2558–2574. <https://doi.org/10.1111/nph.19551>.
- Sur, S., and Ray, R.B. (2020). Bitter Melon (*Momordica Charantia*), a Nutraceutical Approach for Cancer Prevention and Therapy. *Cancers* **12**:2064. <https://doi.org/10.1002/9780470543672.ch2>.
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* **74**:174–183. <https://doi.org/10.1111/tpj.12105>.
- Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *CP in Bioinf.* **25**, UNIT 4.10. <https://doi.org/10.1002/0471250953.bi0410s25>.
- Tian, Y., Li, K., Li, T., Gai, W., Zhou, J., Deng, X.W., Xue, Y., Deng, Y., He, H., and Zhang, X. (2025). The near-complete genome assembly of pickling cucumber and its mutation library illuminate cucumber functional genomics and genetic improvement. *Mol. Plant* **18**:551–554. <https://doi.org/10.1016/j.molp.2025.03.001>.
- Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Tani, N., Miyagi, N., Fukushima, M., Suzuki, S., Tarora, K., Tamaki, M., et al. (2017). Draft genome sequence of bitter melon (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* **24**:51–58. <https://doi.org/10.1093/dnares/dsw047>.
- Villasante, A., Méndez-Lago, M., Abad, J.P., and Montejó de Garcini, E. (2007). The birth of the centromere. *Cell Cycle* **23**:2872–2876. <https://doi.org/10.4161/cc.6.23.5047>.
- Vollger, M.R., Kerpedjiev, P., Phillippy, A.M., and Eichler, E.E. (2022). StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps. *Bioinformatics* **38**:2049–2051. <https://doi.org/10.1093/bioinformatics/btac018>.
- Vollmannová, A., Bojnanská, T., Musilová, J., Lidiková, J., and Cifrová, M. (2024). Quercetin as one of the most abundant represented biological valuable plant components with remarkable chemoprotective effects - A review. *Heliyon* **10**:e33342. <https://doi.org/10.1016/j.heliyon.2024.e33342>.
- Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**:2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>.
- Wang, H.Y., Li, P.F., Wang, Y., Chi, C.Y., Jin, X.X., and Ding, G.H. (2024a). Overexpression of cucumber CYP82D47 enhances resistance to powdery mildew and *Fusarium oxysporum* f. sp. *cucumerinum*. *Funct. Integr. Genomics* **24**:14. <https://doi.org/10.1007/s10142-024-01287-1>.
- Wang, Y., Zhou, F., Li, Y., Yu, X., Wang, Y., Zhao, Q., Feng, X., Chen, J., and Lou, Q. (2024b). Characterization of the CsCENH3 protein and centromeric DNA profiles reveal the structures of centromeres in cucumber. *Hortic. Res.* **11**:uhae127. <https://doi.org/10.1093/hr/uhae127>.
- Wang, S., Wang, H., Jiang, F., Wang, A., Liu, H., Zhao, H., Yang, B., Xu, D., Zhang, Y., and Fan, W. (2022). EndHiC: assemble large contigs into chromosome-level scaffolds using the Hi-C links from contig ends. *BMC Bioinf.* **23**:528. <https://doi.org/10.1186/s12859-022-05087-x>.
- Wang, Z., and Xiang, C. (2013). Genetic mapping of QTLs for horticulture traits in a F2-3 population of bitter melon (*Momordica charantia* L.). *Euphytica* **193**:235–250. <https://doi.org/10.1007/s10681-013-0932-0>.
- Wang, Y., Mostafa, S., Zeng, W., and Jin, B. (2021). Function and Mechanism of Jasmonic Acid in Plant Responses to Abiotic and Biotic Stresses. *IJMS* **22**:8568. <https://doi.org/10.3390/ijms22168568>.
- Wang, D., Li, Y., Wang, H., Xu, Y., Yang, Y., Zhou, Y., Chen, Z., Zhou, Y., Gui, L., Guo, Y., et al. (2023). Boosting wheat functional genomics via an indexed EMS mutant library of KN9204. *Plant Commun.* **4**:100593. <https://doi.org/10.1016/j.xplc.2023.100593>.
- Wlodzimierz, P., Hong, M., and Henderson, I.R. (2023). TRASH: Tandem Repeat Annotation and Structural Hierarchy. *Bioinformatics* **39**:btad308. <https://doi.org/10.1093/bioinformatics/btad308>.
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**:100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
- Xin, H., Liu, X., Chai, S., Yang, X., Li, H., Wang, B., Xu, Y., Lin, S., Zhong, L., Liu, B., et al. (2024). Identification and functional characterization of conserved cis-regulatory elements responsible for early fruit development in cucurbit crops. *Plant Cell* **36**:2272–2288. <https://doi.org/10.1093/plcell/koae064>.
- Xiong, L., Huang, Y., Liu, Z., Li, C., Yu, H., Shahid, M.Q., Lin, Y., Qiao, X., Xiao, J., Gray, J.E., et al. (2022). Small EPIDERMAL PATTERNING FACTOR-LIKE2 peptides regulate awn development in rice. *Plant Physiol.* **190**:516–531. <https://doi.org/10.1093/plphys/kiac278>.
- Yang, J., Weng, Y., Li, H., Kong, Q., Wang, W., Yan, C., and Wang, L. (2022). Epidermal Patterning Factor 2-like (McEPFL2): A Putative Candidate for the Continuous Ridge (cr) Fruit Skin Locus in Bitter Melon (*Momordica charantia* L.). *Genes* **13**:1148. <https://doi.org/10.3390/genes13071148>.
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E.,

- et al.** (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**:650–659. <https://doi.org/10.1093/bioinformatics/bti042>.
- Zeng, Q., Wei, M., Li, S., Wang, H., Mo, C., Yang, L., Li, X., Bie, Z., and Kong, Q.** (2024). Complete genome assembly provides insights into the centromere architecture of pumpkin (*Cucurbita maxima*). *Plant Commun.* **5**:100935. <https://doi.org/10.1016/j.xplc.2024.100935>.
- Zhang, M., Song, M., Davoudi, M., Cheng, F., Yin, J., Zha, G., Yang, Z., Chen, J., and Lou, Q.** (2022). The mutation of C-24 reductase, a key enzyme involved in brassinolide biosynthesis, confers a novel compact plant architecture phenotype to cucumber. *Theor. Appl. Genet.* **135**:2711–2723. <https://doi.org/10.1007/s00122-022-04144-6>.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al.** (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**:R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
- Zhang, Y., Zhao, M., Tan, J., Huang, M., Chu, X., Li, Y., Han, X., Fang, T., Tian, Y., Jarret, R., et al.** (2024). Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat. Genet.* **56**:1750–1761. <https://doi.org/10.1038/s41588-024-01823-6>.
- Zhang, Y., Zhang, Z., Sun, X., Zhu, X., Li, B., Li, J., Guo, H., Chen, C., Pan, Y., Liang, Y., et al.** (2019). Natural alleles of GLA for grain length and awn development were differently domesticated in rice subspecies japonica and indica. *Plant Biotechnol. J.* **17**:1547–1559. <https://doi.org/10.1111/pbi.13080>.
- Zhong, J., Cheng, J., Cui, J., Hu, F., Dong, J., Liu, J., Zou, Y., and Hu, K.** (2022). *MC03g0810*, an Important Candidate Gene Controlling Black Seed Coat Color in Bitter Gourd (*Momordica* spp.). *Front. Plant Sci.* **13**:875631. <https://doi.org/10.3389/fpls.2022.875631>.