Phased Telomere-to-Telomere Reference Genome and Pangenome Reveal an Expansion of Resistance Genes during Apple Domestication

Ying Su[1, 2, †], Xuanwen Yang[2, 3, †], Yuwei Wang[1, †], Jialei Li[1, †], Qiming Long[2], Shuo Cao[2], Xu Wang[2], Zhenya Liu[2], Siyang Huang[2], Zhuyifu Chen[2], Yanling Peng[2], Fan Zhang[2], Hui Xue[2], Xuejing Cao[2], Mengyan Zhang[2], Gulbar Yisilam[1], Zhenzhou Chu[1], Yuan Gao[4], Yongfeng Zhou[2, 5, *], Zhongjie Liu[2, *], Hua Xiao[2, *] and Xinmin Tian[1,6, *]

1 Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Xinjiang, Urumqi 830046, China

2 National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

3 National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Science, Zhengzhou 450009, China

4 Research Institute of Pomology, Chinese Academy of Agricultural Sciences, Key Laboratory of Horticultural Crop Germplasm Resources Utilization, Ministry of Agriculture and Rural Affairs of the People's Republic of China, Xingcheng 125100, China

5 National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

6 Key Laboratory of Ecology of Rare and Endangered Species and Environmental Protection (Ministry of Education) & Guangxi Key Laboratory of Landscape Resources Conservation and Sustainable Utilization in Lijiang River Basin, Guangxi Normal University, Guilin, 541006, China

*Corresponding authors: Xinmin Tian (tianxm06@lzu.edu.cn; X.T.), Hua Xiao (xiaohua01@caas.cn; H.X.), Zhongjie Liu (liuzhongjie@caas.cn; Z.L.), and Yongfeng Zhou (zhouyongfeng@caas.cn; Y.Z)

† These authors contributed equally to this work.

The author responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Xinmin Tian (tianxm06@lzu.edu.cn).

Short title: Expansion of resistance genes during apple domestication

Abstract

The cultivated apple (*Malus domestica* Borkh.) is a cross-pollinated perennial fruit tree of great economic importance. Previous versions of apple reference genomes were unphased, fragmented, and lacked comprehensive insights into the highly heterozygous genome, which impeded genetic studies and breeding programs in apple. In this study, we assembled a haplotype-resolved telomere-to-telomere reference genome for the diploid apple cultivar Golden Delicious. Subsequently, we constructed a pangenome based on twelve assemblies from wild and cultivated apples to investigate different types of resistance gene analogs (RGAs). Our results revealed the dynamics of the gene gain and loss events during apple domestication. Compared with cultivated species, more gene families in wild species were significantly enriched in oxidative phosphorylation, pentose metabolic process, responses to salt, and abscisic acid biosynthesis process. Interestingly, our analyses demonstrated a higher prevalence of RGAs in cultivated apples than their wild relatives, partially attributed to segmental and tandem duplication events in certain RGAs classes. Other types of structural variations, mainly deletions and insertions, have affected the presence and absence of TIR-NB-ARC-LRR (TNL), NB-ARC-LRR (NL), and CC-NB-ARC-LRR (CNL) genes. Additionally, hybridization/introgression from wild species has also contributed to the expansion of resistance genes in domesticated apples. Our haplotype-resolved T2T genome and pangenome provide important resources for genetic studies of apples, emphasizing the need to study the evolutionary mechanisms of resistance genes in apple breeding programs.

Keywords: *Malus*, haplotype-resolved T2T genome, resistance genes, introgression, structural variation, gene duplication

Introduction

The cultivated apple (*Malus domestica* Borkh.) stands as one of the most extensively produced and economically important fruit crops worldwide (Duan et al. 2017). Modern cultivated apples, derived from *Malus sieversii* in the Tian Shan Mountains between 4,000 and 10,000 years ago, have experienced extensive hybridization and introgression with wild crabapple species from Siberia (*Malus baccata* (L.) Borkh.) and Europe (*Malus sylvestris* Mill.) (Cornille et al. 2014; Daccord et al. 2017; Chen et al. 2021; Liao et al. 2021). This process led to a highly divergent genetic architecture, complex genomic features, and high heterozygosity in cultivated apples. The century-old *Malus domestica* cv. Golden Delicious (GD), continues to be popular in global apple production and serves as a foundational ancestor for many modern apple varieties. Previous studies have produced multiple versions of the apple reference genome, including the unphased GD reference genome based on highly homozygous material (GDDH13) (Velasco et al. 2010; Daccord et al. 2017). The genome includes a large number of highly repetitive regions that introduce hundreds or thousands of gaps, creating a significant barrier to achieving a haplotype-resolved complete reference genome (Giani et al. 2020).

Advances in long-read sequencing technologies, such as the Pacific Biosciences high-fidelity (PacBio HiFi) and Oxford Nanopore Technology (ONT) ultra-long platforms, have facilitated the assembly of haplotype-resolved telomere-to-telomere (T2T) genome. The haplotype-resolved T2T genome, which retains allelic sequences to the greatest extent and enhances the accuracy and completeness of assemblies (Yue et al. 2023), is critical for unraveling the full genetic complexity of the apple's highly heterozygous genome. However, a linear reference genome from a single individual is insufficient to represent the entire genetic diversity of a species, especially in light of the diversification and alterations in the genetic structure related to long-term crop domestication (Liu et al. 2020; Yang et al. 2022; Long et al. 2024). To overcome this limitation, pangenome reference has emerged as a promising approach to capture the genetic diversity of crops and their wild relatives (Sherman 2020; Della Coletta et al.

2021; Shang et al. 2022; Tang et al. 2022; Zhou et al. 2022; Chen et al. 2023a; Liao et al. 2023). To date, several wild and cultivated apple chromosome-level genomes have been released, which are important resources for genetic studies (Sun et al. 2020). Consequently, employing a pangenome approach to explore the genetic differences between wild and cultivated apples will provide deeper insights into the genetic mechanisms underlying apple domestication, including disease resistance.

The disease resistance mechanisms became a major focus in breeding and improvement of cultivated apples (Zhang et al. 2021; Karlström et al. 2022). Scab and powdery mildew are the two important fungal threats to apple orchards, causing severe yield and economic losses by damaging leaves and young fruits (Bus et al. 2010). Due to intense selective pressure, plants have developed a variety of mechanisms to recognize and defend against pathogens. Some resistance gene analogs (RGAs) encode proteins containing nucleotide-binding sites (NBS) and leucine-rich repeat sequences (LRR), playing a pivotal role in recognition and resistance across a diverse range of pathogens (Yang et al. 2013; Long et al. 2024). Typically, apple resistance to fungal pathogens originates from single and multigene loci. The presence and distribution of RGAs on apple chromosomes are highly specific, and these RGAs evolve independently within the chromosomes (Perazzolli et al. 2014). It is hypothesized that duplication, divergence, and loss of genome segments lead to changes in the number of RGAs, which in turn promotes species evolution in response to rapidly changing pathogens (Arya et al. 2014; Van et al. 2019). The hybridization and selection contributed to the diversification of nucleotide-binding and leucine-rich repeat (NLR) genes during apple domestication (Perazzolli et al. 2014). Studies have shown that the divergence of resistance genes between wild and cultivated apples means that genes and pathways associated with resistance in apple domestication have been subject to selection (Singh et al. 2021). Additionally, introgression facilitates local adaptation by introducing beneficial alleles from a crop wild relative (CWR) into a cultivated crop, such as enhancing traits such as disease and stress resistance (Xiao et al. 2023). Despite the importance of RGAs, the genetic variations in

resistance genes between wild and cultivated apple varieties have not yet been extensively studied.

In this study, we generated a haplotype-resolved T2T reference genome of the apple variety 'Golden Delicious' (GDT2T) by integrating HiFi, ONT ultra-long, and high-throughput chromosome conformation capture (Hi-C) reads. Our genome assembly exhibits significant improvements in contiguity, completeness, and accuracy when compared to the previously assembled reference genome (GDDH13). Pan-genomic studies have revealed significant structural variation and differences in RGAs between wild and cultivated apple species. Gene duplications, structural variations, and introgression have contributed to the expansion of resistance genes during apple domestication. These extensive genomic variations provide critical resources and insights for accelerating the integration of wild species into apple breeding programs for enhancing disease resistance.

Results

A haplotype-resolved telomere to telomere reference genome for apple (GDT2T)

A total of 27.75 Gb (~42× coverage) PacBio HiFi reads, 11.04 Gb (~17× coverage) ONT ultra-long reads, and 46.23 Gb (~71× coverage) Hi-C reads were generated for GD **(Fig. 1, A and B)** genome assembly. The genomic homozygosity was assessed to be 98.9% using the *k*-mer metric (**Supplementary Fig. S1**). The de novo assembly conducted with Hifiasm produced a contig-level assembly with a contig N50 size of 34.58 Mb. This represents a 55.8-fold improvement compared to the 620 kb of the GDDH13 genome. Subsequently, we employed Juicer and 3D-DNA for scaffold construction and further manually filled gaps to refine the assembly. Finally, the assembled genome sizes for haplotype 1 (GDT2T_hap1) and haplotype 2 (GDT2T_hap2) were 650,775,246 bp and 644,751,625 bp, respectively **(Fig. 1C, Table 1)**. The phasing of these assembled haplotypes was demonstrated in a spectra graph generated by the KAT program (**Supplementary Fig. S2).** The Hi-C contact matrix displayed high consistency across all chromosomes of GDT2T_hap1 and

GDT2T_hap2, validating the accuracy of their ordering and orientation **(Fig. 1D)**. The completeness of the diploid assemblies was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO), with scores indicating 98.7% and 98.6% completeness for GDT2T_hap1 and GDT2T_hap2, respectively. The GC contents of the two haplotypes were 38.08% and 38.05%, separately **(Table 1)**. The *k*-mer analyzer estimated the average quality value (QV) of each chromosome for the two haplotypes to be 61.25 using HiFi reads. Then, we analyzed the phasing accuracy and observed an average haplotype switch error rate of 0.94% and hamming error rate of 1.05%. Genome accuracy was also assessed based on high mapping rates (> 99%) of HiFi reads to each assembly **(Supplementary Table S1)**. All these statistics for evaluation were comparable with recently published complete genomes (Nurk et al. 2022, Chen et al. 2023, Huang et al. 2023, Shang et al. 2023, Shi et al. 2023, Sun et al. 2023). Comparative genome analysis revealed significant synteny among GDT2T_hap1, GDDH13, and GDT2T_hap2 **(Fig. 1C, Supplementary Fig. S3)**. The high degree of collinearity between the genetic distances of the SNP markers and the physical distance on each chromosome further demonstrated the quality of our assembly **(Supplementary Fig. S4)**.

We successfully achieved a phased gap-free reference genome assembly for each haplotype of GD, fulfilling all 3,109 gaps identified in the GDDH13 genome**.** We also corrected previously misassembled regions, notably the start of chromosome 1 and the end of chromosome 16 in GDT2T_hap1, covering regions of 5,416,121 bp (including 111 genes) and 2,722,516 bp (including 63 genes), respectively (**Supplementary Fig. S5, Supplementary Table S2)**. The genes located in misassembled regions were significantly enriched in pathways related to protein deacetylation and cell growth **(Supplementary Table S2).** The authenticity of the inversion between two haplotypes was confirmed by mapping ONT, HiFi, and Hi-C reads to GDT2T_hap1 and by examining Hi-C heatmaps (**Supplementary Fig. S6**). Additionally, we integrated a 2.36 Mb sequence into chromosome 10 (covering 35 genes), including a polyol-related gene (UDP-glucose flavonoid 3-O-glucosyltransferase 7, named *GT7*)

that is potentially critical for sugar decomposition in strawberry (*Fragaria* × *ananassa*) (Griesser et al. 2008) (**Fig. 1E, Supplementary Fig. S7, Supplementary Table S3)**. A comprehensive and accurate assembly of the complete diploid genome represents a major advance in apple genomic research.

Gene and TE annotations of GDT2T

For gene content assessment, we identified 44,375 and 44,304 protein-coding genes in GDT2T_hap1 and GDT2T_hap2, capturing 96.6% and 97.0% of a BUSCO reference gene set **(Table 1)**. For GDT2T_hap1 and GDT2T_hap2, there are 63,512 and 63,437 transcripts respectively, with an average of 1.43 splice variants per gene. Of these, 36,443 (94.51%) in GDT2T_hap1 and 34,221 (94.62%) in GDT2T_hap2 were functionally annotated using the comprehensive Pfam database. Furthermore, a total of 1,396 genes were specific to GDT2T_hap1, and 1,306 genes were unique to GDT2T_hap2. These distinctions were the results of ancestral apple hybridization followed by subsequent clonal propagation. Functional enrichment analysis based on Gene Ontology (GO) annotation indicated these unique genes were significantly enriched in the triterpenoid biosynthetic process (**Supplementary Table S4**).

We utilized the Extensive de novo TE Annotator (EDTA) to generate a high-quality repetitive sequence library (Xie et al. 2019). A total of 866,404 and 899,134 transposable elements (TEs) were identified in GDT2T_hap1 and GDT2T_hap2, accounting for 61.5% and 61.47% of the two haplotypes, respectively (**Supplementary Table S5)**. In the GDT2T_hap1, long terminal repeat retrotransposons (LTR-RTs) emerged as the most prevalent TE class, with *Gypsy* and *Copia* LTR-RTs comprising 18.89% and 10.36% of the genome, respectively.

To define the telomeric regions in GD's two haplotypes, the most prevalent telomere repeat unit (CCCATTT at the 5' end and TTTAGGG at the 3' end) was identified, which was present in nearly all chromosomes (Fajkus et al. 2005). We predicted 30 telomeres in both haplotypes. In GDT2T_hap1, the longest telomere was located at the end of chromosome 15 (12.87 kb) and consists of 1,839 repeats. Conversely, chromosome 14 exhibited the shortest telomere at 609 bp with 87 repeats

(Supplementary Table S6). Regarding centromeric regions, the identification of an optimal centromeric repeat unit was challenging due to the presence of dispersed fragments of varying lengths within chromosomes. Typically, LTR-RTs are enriched in plant centromeres often mixed with tandem repeats (Fernandes et al. 2019; Song et al. 2021). Over 90% of rice centromeric repeats are *Gypsy*, which complicates complete centromere assembly (Song et al. 2021). In the apple genome, Gypsy elements were notably enriched in the centromeric regions of both haplotypes, while LTR (*Copia*), DNA (*MULE-MuDR*), and RC (*Helitron*) TEs were absent **(Supplementary Fig. S8, Supplementary Table S6).** Additionally, we confirmed the position of the centromere using previously published HODOR sequences (Daccord et al. 2017). By examining the distribution regions of various TEs and HODOR sequences, we delineated the centromeric ranges of all chromosomes in both haplotypes, identifying regions that closely align with previous findings in 'HFTH1' (Zhang et al. 2019).

Hemizygosity within the diploid GD genome

Apple domestication is driven by the hybridization of different wild species and the clonal propagation by artificial selection of genotypes with desired traits. This process likely contributes significantly to phenotypic variation through heterozygous genomic regions (Schnable and Springer 2013). In this study, we assessed the impact of heterozygous structural variations (hSVs) on these genes. Using Sniffles (v2.0.7), we remapped HiFi reads to the GDT2T_hap1 reference genome and identified 56,852 hSVs (> 50 bp), comprising 28,720 deletions, 23,799 insertions, 40 duplications, 147 inversions, and 4,146 translocations (**Supplementary Table S7**). Furthermore, 3,323 hemizygous genes were found in the apple genome, accounting for 7.5% of all annotated protein-coding genes. Coverage values in hemizygous regions, as verified using the Integrated Genomics Viewer (IGV), showed a reduction to half of their original levels (**Supplementary Fig. S9**). GO enrichment analysis showed that the

functions of hemizygous genes are related to important biological processes, including protein phosphorylation and defense response process **(Fig. 1F).**

Pangenome and structural variation among wild and cultivated apples

To build a comprehensive gene pool for apple varieties, we performed pangenome analysis using seven apple accessions (Chen et al. 2019; Zhang et al. 2019; Sun et al. 2020; Khan et al. 2022; Li et al. 2022). These included four accessions from *Malus domestica* (GD, Gala, Honeycrisp, and HFTH1), three wild apple accessions (including one *Malus baccata* (Mba), one *Malus sieversii* (Msi), and one *Malus sylvestris* (Msy)), representing diversity of the *Malus* genus (**Supplementary Fig. S10, Supplementary Table S8).** Orthologs analysis categorized all the protein-coding genes from seven apple genomes into 44,989 pan-genomic clusters (**Fig. 2A**). The gene family counts increased rapidly, stabilizing as more genomes were included, while the counts of core genes gradually declined (**Fig. 2B**). Of the total gene sets, 34.41% were core genes (present in seven accessions), 63.43% were dispensable genes (present in five accessions), and 1.15% were private genes (present in one accession) (**Fig. 2C, Supplementary Table S9**). Core genes exhibited significantly lower pairwise d_N/d_S ratios than private genes, indicating their functional conservation (**Fig. 2D**). Additionally, 90.51% of the core genes were found to possess InterPro domains, approximately 1.7 times more than private genes **(Fig. 2E)**. These findings collectively emphasized the immense value of pangenome studies in exploring and utilizing diverse gene pools, which is crucial for advancing apple biology and breeding programs.

To investigate alterations in genetic composition throughout the process of apple domestication, we examined genomic differences between wild and cultivated apples based on the pangenome (n=7). We found that 609 gene families had a significantly higher gene count in cultivated species compared to wild species. GO enrichment analysis revealed that these genes were mainly enriched in the regulation of cellular respiration, energy derivation, as well as regulation of cellular component size

biosynthetic process. Conversely, the 338 gene families more abundant in wild species were involved in oxidative phosphorylation, pentose metabolic process, responses to salt, and abscisic acid biosynthesis process (**Fig. 2, F and G**). These findings highlight significant differences in secondary metabolite synthesis between wild and cultivated apples.

To further understand genomic divergence within *Malus*, we conducted whole-genome alignments of eleven genomes using GDT2T_hap1 as the reference. Finally, 28,219 high-confidence structural variations (SVs) (> 50 bp) were identified, including 13,159 insertions, 9,406 deletions, 3,246 inversions, and 2,408 duplications **(Supplementary Table S10**). These SVs were mostly scattered across the 17 chromosomes, with insertions and deletions being less frequent in centromeric regions, whereas duplications were more prevalent (**Supplementary Fig. S11)**. Notably, a fixed large inversion (~523 kb) on chromosome 11 was identified in the genomes of all wild and domesticated apples studied (except Gala_hap2) (**Supplementary Fig. S12, Supplementary Table S11**). Within this inversion region, 29 genes were identified, including a cytochrome P450 protein potentially involved in environmental stress response mechanisms, such as detoxification of plant defensive compounds or insecticides (**Supplementary Table S12**) (Heitz et al. 2012). The extensive variation among wild and cultivated apple species provides access to further harness the genetic diversity of distantly related wild apples in genome-driven breeding.

RGAs in the wild and cultivated apples

Plant resistance to various pathogens can be controlled by a variety of plant resistance (R) genes that exhibit different evolutionary patterns across species. In our study, we systematically searched for plant RGAs across twelve apple genomes, employing RGAugury (Li et al. 2016), NLR-annotator (Steuernagel et al. 2020) and DeepLRR (Liu et al. 2022). These RGAs were classified into eleven families based on protein domains, namely CN, CNL, NBS, NL, TN, TNL, TX, RLK, RLP, RPW8-X, and TM-CC (see the method for specific classifications). Our analysis revealed a range of

1,012 to 1,309 RGAs in each of the twelve genomes from both wild and cultivated apples (**Supplementary Fig. S13**). Notably, cultivated apples exhibited a significantly higher number of RGAs compared to their wild counterparts **(Fig. 3, A and B)**. Some types of RGAs, such as TNL, were more prevalent in cultivated apples, averaging a 29.8% increase over wild apples **(Fig. 3A, Supplementary Table S13)**. On average, TM-CC was the most abundant, accounting for 25.9% of all RGAs, followed by RLK (20.5%) and TNL (13.61%). The least abundant were RPW8-X (1.15%), CN (1.25%), and TN (2.18%) (**Supplementary Table S13**). Furthermore, we demonstrated the distribution characteristics of RGAs on each chromosome in both wild and cultivated apples (**Fig. 3C, Supplementary Fig. S14)**. On average, chromosomes 2, 5, 10, 15, and 11 harbored 9.2%, 9.1%, 9.0%, 7.9%, and 7.9% of the anchored RGAs, respectively. In contrast, chromosomes 6, 12, 13, 14, and 16 contained only 4.4%, 4.2%, 2.9%, 4.2%, and 3.1% RGAs, respectively (**Supplementary Table S14**). The distribution pattern was consistent with that reported in previous studies (Singh et al. 2021). Certain specific RGAs have been validated, such as the *MdRNL6*, known for its contribution to broad-spectrum fungal resistance in apples (Perazzolli et al. 2014) **(Supplementary Table S15)**.

Generally, the NLR gene arrangement patterns were clustered or head-to-head (paired) (Van et al. 2019). Compared to wild apples, cultivated varieties exhibited a higher number of both NLR gene clusters and pairs **(Fig. 3B, Supplementary Table S16)**. Furthermore, integrating all RGAs identified per genome with pan-genomic results revealed that 58.1% of these RGAs were core genes, 38.8% were dispensable genes, and 3.1% were private genes (**Supplementary Table S13**). Overall, genetic recombination resulting from ancestral hybridization has played a significant role in diversifying the genetic structures of RGAs in both wild and domesticated apple varieties. To study the dynamic patterns of NLR genes in wild and cultivated apples and their relationship with gene duplication, we constructed a phylogenetic tree based on the protein sequences of different types of NLR genes. It has been observed that gene duplication was more prevalent in cultivated species than in their wild relatives

within the same phylogenetic branch, suggesting that gene duplication contributes significantly to the enhanced accumulation of NLR genes in cultivated apples (**Supplementary Fig. S15**). Furthermore, we analyzed the duplication types of RGAs in both cultivated and wild apples based on the location of RGAs on the genome. An average of approximately 33.4% of RGAs underwent whole-genome/segmental duplications (WGD), while 25.3% of RGAs experienced dispersed duplication events **(Fig. 3D)**. Tandem and proximal duplication events accounted for 24.4% and 7.9% of all duplication events, respectively. Additionally, we observed differences in RGA classes associated with various duplication events. For example, most CNL genes were preserved through tandem duplications, whereas RLK genes predominantly underwent segmental duplications **(Fig. 3D)**. Comparing the duplication types of RGAs between cultivated and wild apples, we observed that in cultivated apples, TNL, TX, and RLP genes showed a significant increase in proximal duplication events. NBS and TNL genes exhibited a significant increase in dispersed duplication events, while RLK and CN genes were significantly amplified through segmental and tandem duplications, respectively. In contrast, a significant increase in the number of TM-CC was observed exclusively in wild apples when in a singleton state. Moreover, SVs, particularly deletions and insertions, were identified as key mechanisms impacting RGAs. On average, TNL, CNL, NL, and TM-CC were most affected by SVs, with proportions of 41%, 39%, 37.2%, and 37.1%, respectively, whereas TX genes were the least affected, at 21.3% (**Fig. 3E)**.

Evolution and introgression of RGAs between cultivated and wild apples

To further investigate the differences in RGAs content between wild and domesticated apples, RGAs from twelve genomes were classified into 1,182 families. Certain RGAs families were present in cultivated apples but absent in their wild counterparts. For instance, the OG0000045 gene family experienced an expansion in the Msy and all cultivated genomes, but it was absent in Msi and Mba (**Fig. 3F)**. This suggested that RGAs of cultivars may have originated from Msy, highlighting a

selective genetic adaptation in cultivated species. Moreover, the RGAs have expanded in cultivars; for instance, the resistance gene N, which mediates resistance to the tobacco mosaic virus (TMV), is present in 2-3 copies in cultivated species compared to only one copy in the two wild species **(Supplementary Table S17)**.

Previous studies proposed that *Malus domestica* originated from Msi and later incorporated substantial genetic contributions from Msy (Cornille et al. 2012). Additionally, the possibility of genetic introgression from other wild *Malus* species implies a complex domestication history. To uncover the origins of RGAs in cultivated apples, we performed homologous sequence comparisons with wild relatives, aiming to identify the ancestral species contributing most significantly to the RGAs content in cultivated species. Comparative analysis of the four genomes revealed 474 (GD, Msi, Msy, and Mba) and 479 (Gala, Msi, Msy, and Mba) common RGAs **(Fig. 4A)**.

In the scenarios lacking common RGAs, GD exhibited a 39.8% overlap of RGAs with Msi, which was slightly lower than the 40.0% shared with Msy, and higher than the 32.3% overlap with Mba. It has been suggested that the predominant contribution of RGAs from Msi and Msy in their hybridization with cultivated apple species **(Fig. 4B)**. Gala showed a higher proportion of RGAs (39.5%) with Msy compared to 38.9% with Msi and 31.5% with Mba, suggesting a more significant exchange of RGAs with Msy.

Mba, known for disease resistance and cold tolerance, is frequently hybridized with cultivars to enhance the disease resistance of fruit trees (Cornille et al. 2012; Chen et al. 2019). However, the precise integration of the resistance genes into cultivars and their chromosomal locations in the introgressed segments had remained unidentified. In our study, we selected genomes from three wild (Msi, Msy, and Mba) species and the cultivated apple (GD) for collinear sequence alignment. A genomic fragment (Chr02: 3,001,525-3,195,910bp) in the GDT2T_hap1 genome, closely matching Mba and distinct from Msi and Msy, suggested potential introgression from Mba **(Fig. 4C)**. To verify the hypothetical introgression event, we calculated the genetic distances

between the syntenic regions of Msi, Msy, and Mba relative to the GD genome. Consistent with previous findings, there were significant divergences between the two haplotypes of GD across chromosomes corresponding to the hybrid origin of the wild-species alleles (Sun et al. 2020) **(Fig. 4D)**. However, some chromosome regions of Mba displayed a closer genetic divergence to the GDT2T_hap1 genome. Specifically, within a 100 kb region, GD showed significant genetic divergence between its two haplotypes. However, the genetic distance between GDT2T_hap1 and Mba reduced to 0.67. This region also showed a large genetic divergence between GD and the other two wild species **(Fig. 4D, Supplementary Table S18)**, suggesting potential introgression from Mba into cultivated apples and consequent RGAs expansion. In this introgressed fragment, we identified two genes that specifically introgressed from Mba into GD, as the TNL and NL types (*MDgdA02G000320* and *MDgdA02G000324*), but 10 amino acids were altered in both of them during long-term introgression. By comparing these two putative introgressed genes with grape homologous gene families, we found that one of the homologous genes (*MDgdA02G000320*) has the closest phylogenetic relationship to genes that have been verified to be associated with downy mildew in grapes (Feechan et al. 2013) (**Supplementary Table S19, Supplementary Fig. S16)**. Furthermore, 10 amino acid alterations were identified in the *MDgdA02G000320*. However, functional annotations indicated that 91.7% of the SNP mutations in the gene had moderate to minor effects on proteins, suggesting that this gene may have played similar roles in apples (**Supplementary Table S20)**.

Discussion

Modern apple growers focus on producing cultivars that meet the specific needs of the apple industry, such as high nutritional value, disease resistance, and ecological adaptability (Duan et al. 2017; Zheng et al. 2020; Liao et al. 2021; Lin et al. 2023). Due to the long juvenile period of apple trees, traditional breeding techniques such as repeated backcrossing are extremely time-consuming (Luo et al. 2020), which

emphasizes the need for effective breeding strategies for apples (O'Rourke 2021). In our research, we provide a haplotype-resolved reference genome for apples. Our analyses revealed the expansion of RGAs in cultivated apples by employing pangenome analysis, providing insights into changes in gene function during domestication and accelerating the integration of wild species into apple breeding.

A haplotype-resolved T2T reference genome for apple genetics and breeding

At the molecular level, a high-quality reference genome is indispensable for uncovering traits and facilitating genetic enhancement (Lin et al. 2023). Despite the release and subsequent updates of the Golden Delicious genome since 2010, limitations in sequencing technology and assembly algorithms have led to incomplete assemblies (Velasco et al. 2010). The development of third-generation sequencing technology, especially the combination of high-continuity ONT ultra-long sequencing and high-accuracy HiFi sequencing, has overcome the assembly problems of centromeres and highly repetitive regions. In this study, we have successfully constructed a haplotype-resolved T2T reference genome, correcting many misassembles present in the GDDH13 and filling all chromosomal gaps. This enhanced assembly enabled the identification of several genes not present in the earlier version, which are significantly enriched in protein deacetylation and cellular growth processes. Given the long history of hybridization and clonal reproduction of apples as perennial crops, our investigation uncovered the presence of hemizygous genes within the GD genome (Zhou et al. 2019; Sun et al. 2020). The results indicate that 7.5% of the genes exhibit a hemizygous state, and were significantly enriched in important biological processes, such as protein phosphorylation, defense response, and phosphoenolpyruvate transport, which may lead to changes in apple agronomic traits or resistance. This ratio is lower than the 15.5% reported by Zhou et al. (2019) in grapes, but it exceeds the results observed in several potato accessions (Tang et al. 2022). Overall, the completed genome assembly stands as a foundational reference for

450 further apple research and offers valuable genetic insights for evolutionary studies and
451 advancements in apple breeding.
452
453 **Pangenome reveals gene gain/loss during apple domestication**
454 Domestication can modify the phenotype and genome of various crops through
455 selection (Liao et al. 2023). Comprehending the population genetics of crop
456 domestication is pivotal for advancing genetics (Varshney et al. 2020; Liao et al.
457 2023). The pangenome is a powerful method for uncovering gene content alterations
458 throughout the domestication of apples and for identifying structural variations with
459 reduced reference bias (Li et al. 2022a; Li et al. 2023; Wang et al. 2023). Our
460 pangenome analyses in this study have shed light on the impact of domestication on
461 gene content, with effects predominantly centered around cellular respiration, energy
462 procurement mechanisms, and various secondary metabolic pathways. Compared with
463 cultivated species, expanded genes in wild species enriched in metabolite synthesis
464 pathways, echoing recent observations by Wang et al. (2023). Despite the varying
465 completeness of genomes in our constructed pangenome, we endeavor to minimize
466 the bias from incomplete genomes. For a more accurate exploration of gene pool
467 diversity between wild and cultivated apples, future pangenome construction should
468 prioritize incorporating a broader range of more complete genomes. In a deeper
469 exploration of genomic architecture, we found a large inversion fixed on chromosome
470 11 in the wild apples. This implies that in the process of domestication, hybridization
471 events and subsequent recombination processes can engender SVs (Baurens et al.
472 2019). Such SVs hold the potential to drive both the loss and expansion of resistance
473 genes within both wild and domesticated populations, resulting in changes in
474 population adaptability to pathogens and environments (Long et al. 2024).
475 In recent years, studying the disease resistance mechanism of wild and cultivated
476 germplasm has become crucial for apple breeding and improvement (Bus et al., 2011;
477 Caffier et al., 2016). In this study, the numbers of RGAs were significantly higher in
478 cultivated apples than in wild apples. However, among eleven different types of

RGAs, only five categories showed a significant increase in cultivars compared to wild species. These results can be attributed to the unique naturally controlled habitats and disease pressures of wild and domesticated plants, where cultivation practices favor the emergence of novel host-specific pathogens and their strains, resulting in human-driven selection for an increased number of resistance genes (McDonald and Stukenbrock 2016). The non-random distribution of various RGA classes across chromosomes appears to be an outcome of the domestication process in perennial species, which enhances selection in some genomic regions associated with specific diseases and environmental adaptations (Li et al. 2011). We discovered that most RGAs were amplified through segmental, dispersed, and tandem duplication events. This pattern is consistent with the notion that disease and selection pressures have driven the expansion of NLR genes through both proximal and tandem duplication events (Zhang et al. 2019; Sun et al. 2020).

Introgression from wild apples empowers disease resistance in apple breeding

It is well-known that the domestication of apples has been largely driven by interspecific hybridization (Lee et al. 2003). In previous studies, several disease-resistance genes implemented in commercial apple breeding were introduced from wild Malus species (Peil et al. 2007). In our study, the RGAs of cultivated apples (GD and Gala) were derived from Msy and Msi in almost equal proportions, indicating that RGAs were retained during hybridization in these two wild ancestors. We also observed a few shared RGAs specific to Mba with GD and Gala. Therefore, we used Mba in the comparative analysis to explore whether the RGAs were introduced into the cultivars (GD and Gala) from Mba. Gene introgression frequently occurs in perennial crops like apples, driven by short divergence times and incomplete reproductive isolation between modern crops and their wild relatives (Wang et al., 2023; Chen et al. 2021). Through comparative genomics and genetic distance analysis between GD and wild apple species, we identified a genomic segment containing two resistance genes in the GDT2T_hap1 that resulted from Mba introgression. Sequence

analysis showed that one of the introgressed genes showed 98.6% sequence similarity between wild and cultivated varieties, the significant similarity confirms the introgression event. Notably, the homolog of the TNL type confers resistance to powdery mildew and downy mildew in grapes, demonstrating that the introgression of resistance genes may enhance the disease resistance of cultivated varieties (Feechan et al. 2013). Overall, incorporating wild apple genetic resources into apple breeding programs to select disease- and cold-resistant varieties may be relevant to sustainable apple production (Morales-Cruz et al. 2023). Moreover, by combining rootstocks with grafted selections, we can pyramid several disease-resistance alleles from different sources to create elite breeding parental lines (Luo et al. 2020).

Materials and Methods

Plant materials and sample collection

Fresh and healthy leaf tissue was collected from *Malus domestica* cv. 'Golden Delicious' apple trees, in Wuwei, Gansu Province, China, and immediately preserved using liquid nitrogen. These materials were packaged for HiFi, ONT ultra-long, and Hi-C sequencing, respectively, and subsequently submitted to the company for genomic DNA extraction and library preparation.

Library preparation and DNA sequencing

High-molecular-weight genomic DNAs were isolated using the DNeasy Plant Mini kit (Qiagen) following the Isolation of high-molecular-weight genomic DNAs using the DNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions. For PacBio HiFi sequencing, a single-molecule real-time cell was sequenced on the PacBio Sequel II platform using CCS with default parameters. For ONT ultra-long sequencing, the library was prepared with the Oxford Nanopore SQK-LSK109 kit and then sequenced on the PromethION ONT platform. Hi-C libraries were generated by digesting samples with the DpnII restriction enzyme and constructing them using a standard Hi-C protocol. Subsequently, these Hi-C libraries were sequenced on the Illumina HiSeq X Ten platform.

Genome assembly and quality assessment

GenomeScope2.0 (Ranallo-Benavidez et al. 2020) was utilized to estimate the genome's heterozygosity through a *k*-mer-based approach. The genome was de novo assembled using HiFi reads and Hi-C reads in combination with Hifiasm (v0.16) (Cheng et al. 2021). The initial output of Hifiasm produced a pair of components named p_ctg, including the phased haplotypes. Subsequently, two pseudochromosomes were constructed, validated, grouped, and anchored with Hi-C reads using Juicer (v1.5) (Durand et al. 2016) and 3D-DNA (v180922) (Dudchenko et al. 2017). An interactive heatmap of the two haplotypes was manually edited using Juicebox (v1.11.08) (Durand et al. 2016). The gap information of pseudochromosomes was detected with a Python script (getgaps.py). The ONT ultra-long reads were assembled by NextDenovo (v2.4.0). We manually adjusted the assembly by assembled contigs as well as ONT reads, which were mapped to the genome assembly by Minimap2 (v2.24) (Li 2018) and visualized in Integrated Genomics Viewer (IGV v2.12.3) software to observe whether the gap regions were supported by reads (Shi et al. 2023). The most consistent sequence was applied to manually replace the corresponding gap-joining sequence. Eventually, two gap-free haplotypes of the GD assembly of all 17 chromosomes were obtained.

The haplotype-resolved T2T genome of GD was validated using the KAT program (Mapleson et al. 2017) confirming the successful assembly of both haplotypes. Genome quality and completeness were assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.2.2) (Simão et al. 2015) with the 'eudicots_odb10' database. To detect switch error and hamming error in our phased-genome assembly, we used a pipeline 'calc_switchErr' (https://github.com/tangerzhang/calc_switchErr), relying on a 'true' phased SNP dataset generated using HiFi and Hi-C reads. N50 was used to assess the integrity and continuity of genome assembly, respectively. The raw reads were mapped to the final assembly to estimate the mapping rate and used SAMtools (v1.9) (Danecek et al.

2021) to calculate. Hi-C data were visualized and checked using Juicebox. We collected published SNP map information (Di Pierro et al. 2016). The genetic map was anchored to the syntenic information of the GDT2T_hap1 genome using ALLMAPS (Tang et al. 2015).

Annotation of genes and transposable elements

RNA-sequencing data (SRR10821819, SRR10821824, SRR13867877, SRR12533153, SRR14225143, SRR9887260) were downloaded from the National Center for Biotechnology Information (NCBI) and encompassed leaf, root, and fruit tissues. The raw reads underwent quality filtering using Trimmomatic (v0.39) (Bolger et al. 2014). Transcriptome assembly was performed using HISAT2 (v2.2.1) (Kim et al. 2019) and StringTie (v2.1.7) (Pertea et al. 2015). For gene annotation, we followed the pipeline in the Genome-Wide-Annotation-Pipeline (https://github.com/unavailable-2374/Genome-Wide-Annotation-Pipeline). For gene functional annotation, InterProScan (v5.29) was applied to predict potential protein domains, based on sequence signatures. The transposable elements were identified by the Extensive De-Novo TE Annotator (EDTA, v2.0.0) with default parameters (Ou et al. 2019).

The identification of telomeres and centromeres

The telomere repeat units were explored by using the Telomere Identification toolkit (TIDK v0.2.0) (https://github.com/tolkit/telomeric-identifier) with the 'tidk explore' function. The entire genome was then further searched using the 'tidk search' function. Telomere statistics were generated based on TIDK plots. For centromere localization, we roughly determined their locations using the abundance of different types of TE across chromosomes. We downloaded HODOR repeat sequences and compared their genomic distribution between two haplotypes to validate the accuracy of centromere boundaries (Daccord et al. 2017). Data analysis and visualization were completed using IGV (v2.12.3).

Genome comparison and synteny analysis

Whole-genome alignment was conducted between the assembled sequences (GDT2T_hap1 and GDT2T_hap2) and GDDH13 using MUMmer (v4.0.0) with parameters 'NUCmer -c 90 -l 40' (Marçais et al. 2018). The alignment blocks were then subjected to a filtration process using 'delta-filter' in the one-to-one alignment mode (-1). Detailed comparison results were extracted using the parameters 'show-coords -T -q -H'. Meanwhile, Synteny and Rearrangement Identifier (SyRI) (Goel et al. 2019) was used to identify collinear orthologs and structural rearrangements. Graphical representations were produced using Plotsr (Goel and Schneeberger 2022).

To detect structural variants (SVs) in the seven accessions, long reads were mapped to the GDT2T_hap1 genome via Minimap2 (v2.24) with the parameters 'minimap2 -ax map-pb/minimap2 -ax map-hifi'. Variant calling was performed using four callers: Sniffles (v2.0.7) (Sedlazeck et al. 2018), SVIM (v1.4.2) (Heller and Vingron 2019), CuteSV (v1.0.13) (Jiang et al. 2020), and PBSV (v2.8.0) (https://github.com/PacificBiosciences/pbsv) all set to default parameters. Merging of all SVs from different callers was conducted using SURVIVOR (v1.0.6) (Jeffares et al. 2017). Finally, BCFtools (v1.9) (http://github.com/samtools/bcftools) was utilized to filter SVs across all samples, focusing on variants.

Identification of specific genes and hemizygous genes

We used DIAMOND (v0.9.24) (Buchfink et al. 2015) software to perform all-vs-all sequence alignment on the input sequences. The parameter was 'diamond blastp -d $db -q $data -k 1 -e 0.00001'. To identify the hemizygous genes in GD, we employed Minimap2 (v2.24) to map HiFi reads to the GDT2T_hap1 genome. Subsequently, Sniffles (v2.0.7) with default parameters (Sedlazeck et al. 2018) was utilized for detecting variations. Filter using the following criteria: length $\geq$ 50 bp, quality $\geq$ 60, PASS and PRECISE. These results were consolidated into heterozygous SVs. Genes

entirely overlapping with these deletions were classified as hemizygous genes (Zhou et al. 2019). To understand the functionalities of these hemizygous genes, we conducted a GO enrichment analysis and visualized it (Shi et al. 2023).

Analyses of the pan-genome

For the pan-genomic analyses, ten genomes along with their sequencing reads were obtained from public databases (**Supplementary Table S8**). Within these genomes, *Malus baccata* was aligned and oriented to chromosomes utilizing the reference-guided software RagTag (Alonge et al. 2022), with GDDH13 serving as the reference genome. We used liftoff to update the annotation files of all genomes to ensure the same annotation process (Shumate and Salzberg 2021). We used GTX to call SNPs from the resequencing data of 38 samples (v2.2.1) (http://www.gtxlab.com/product/cat) with GD_hap1 as the reference, and then PLINK (v1.90b6.21) (Purcell et al. 2007) was used for principal component analysis (PCA) analysis (Cornille et al. 2012). Gene family analysis was performed using OrthoFinder (v2.5.2) (Emms and Kelly 2019) based on the longest proteins of the seven accessions. Single-copy orthologous genes were extracted and aligned, and then RAxML was used to construct a phylogenetic tree with pear as the outgroup (Sun et al. 2023). The identified gene clusters were categorized into three groups: core (present in seven accessions), dispensable genes (present in five accessions), and private gene clusters (present in one accession). Syntenic gene pairs were identified using MCscanX (Wang et al. 2012), and Perl scripts (add_ka_and_ks_to_collinearity.pl) utilized for calculating the synonymous (d_S) and nonsynonymous (d_N) substitution rates within the gene clusters, determining the d_N/d_S ratios.

Identification and classification of RGAs

NLR-annotator (v2.1) was used to detect genomic fragments containing putative nucleotide-binding domains and leucine-rich repeat (NLR) genes in each genome (Steuernagel et al. 2020). Subsequently, the RGAugury (v2.2) (Li et al. 2016) pipeline

classified putative nucleotide-binding site (NB-ARC) domain-encoding genes into various subgroups based on domain and motif structures. These subgroups included NBS (NB-ARC), TN (Toll/interleukin-1 receptor (TIR) and NB-ARC), CN (coiled-coil (CC) and NB-ARC), CNL (CC, NB-ARC and LRR), NL (NB-ARC and leucine-rich repeat (LRR)), TNL (TIR, NB-ARC and LRR), TX (TIR-X), RPW8-X (RPW8-like), TM-CC (transmembrane and coiled-coil) (Van de Weyer et al. 2019). Notably, the presence of only LRR or CC motifs was not considered sufficient for NLR identification. Lastly, DeepLRR was utilized to detect leucine-rich repeats containing RLKs (LRR-RLKs), and leucine-rich repeats containing RLPs (LRR-RLPs) (Liu et al. 2022). These identified RGAs were subsequently merged and their distribution across chromosomes was visualized using RIdeogram (Hao et al. 2020). Paired NLR genes were defined as two NLR genes separated by no more than two non-NLR genes. If three or more genes were 'paired', they were classified as a gene cluster. A python script was then employed to count the number of clustered and paired NLRs in each genome. Based on the GO database (http://geneontology.org/), we conducted the background GO-Terms using whole genome proteins. The GO enrichment analysis was performed by clusterProfiler (v4.0) (Wu et al. 2021) and using a test threshold of FDR < 0.05. All *P*-values obtained by Student's t tests (one-tailed) are indicated.

The impact of duplication events and structural variation on RGAs

We utilized the 'intersect' command embedded in BEDTools (v2.29.2) to determine the impact of SVs on RGAs in each genome by comparing their positions with SVs (Quinlan and Hall 2010). We used 'duplicate_gene_classifier' in McscanX software to detect duplication events of different RGAs. We applied the Mann-Whitney U test to evaluate statistically significant differences in the numbers of gene families between wild and cultivated apple species. Protein sequences from the RGAs were aligned using MUSCLE (v3.8.31) (Edgar 2004) with the default parameter. Alignment results were then fed to IQ-TREE (v1.6.12) (Nguyen et al. 2015) for

phylogenetic tree construction. The SNPs on homologous genes were annotated using snpEff (v5.2) (Cingolani et al. 2012) with default parameters, based on the gene annotation for the GDT2T_hap1 genome.

Genetic contribution of wild progenitors to the cultivar apple

Genes of the merged two haplotypes were filtered using CD-HIT (v4.8.1) (option: -c 0.9) (Li and Godzik 2006), followed by the analysis of homologous genes between wild and cultivated species using MCScanX. Venn diagrams were generated via the 'ggvenn' (R package) for visual representation. We performed genome-wide comparisons for each genome using specified parameters 'nucmer -g 5000 -b 1000 -L 10000 -t 15'. The collinearity of six haplotypes was illustrated using Plotsr. To infer the genetic contribution of the three wild ancestors to GDT2T_hap1, we projected the consensus genome fragment of all wild species onto GDT2T_hap1 and used distmat (https://www.bioinformatics.nl/cgi-bin/emboss/distmat) to calculate genetic divergence. Subsequently, a Jukes-Cantor correction was applied to each non-overlapping 50kb window on the GDT2T_hap1 and GDT2T_hap2. The genetic origins of each window of cultivated apples were matched to Msi, Msy, and Mba based on lower variance values. Sequence alignments of candidate proteins were conducted using ClustalW in MEGA11 (Tamura et al. 2021), and sequence visualization was accomplished using GeneDoc (vgd322700) (KB 1997).

Accession Numbers

Sequence data and gene name from this article can be found in the GDR (https://www.rosaceae.org/species/malus/all).

Supplementary Data

Supplementary Figure S1. Assessment of 'Golden Delicious' genome size and heterozygosity by *k*-mer.

Supplementary Figure S2. *K*-mer analysis shows comparisons of HiFi reads to the genome with KAT.

Supplementary Figure S3. A comparison of genomic collinearity between the two haplotypes of GD and GDDH13, respectively.

Supplementary Figure S4. Comparative analysis of the physical location of GDT2T_hap1 and the genetic distance of SNP markers in four previously published genetic maps.

Supplementary Figure S5. Verification of real reads in the inversion region between GDDH13 and GDT2T_hap1.

Supplementary Figure S6. Verification of real reads in the inversion region between GDT2T_hap1 and GDT2T_hap2.

Supplementary Figure S7. Sequence alignment of *GT7* homologous genes: includes '*GT7*' (strawberry), '*XP_008364526.1*' (*Malus*), '*GDT2T_hap110g00922*' (GD).

Supplementary Figure S8. Visualization of transposable elements (TEs) and HODOR repeat sequences across the whole genome using IGV.

Supplementary Figure S9. A demonstration of the hemizygous gene of GD supported by homozygosity, coverage, and the breakpoints of HiFi reads.

Supplementary Figure S10. Selected apple genome information for pangenome studies.

Supplementary Figure S11. The landscape of structural variations among wild and cultivated apples.

Supplementary Figure S12. Inversion map of wild and cultivated apple genomes.

Supplementary Figure S13. Comparison of resistance gene analogs (RGAs) among different genomes.

Supplementary Figure S14. The distribution plots of RGAs across different chromosomes.

Supplementary Figure S15. Evolutionary trees of the five classes of RGAs between wild and cultivated apples.

Supplementary Figure S16. Comparison of the two introgressed genes from *Malus baccata* to GD.

Supplementary Table S1. Summary of sequencing data.

Supplementary Table S2. Functional annotation of genes in the corrected regions of the GD genome.

Supplementary Table S3. Functional annotation of genes in the newly integrated region on chromosome 10 of the GDT2T_hap1 genome.

Supplementary Table S4. GO functional annotation of all private genes identified between the two haplotypes of GD.

Supplementary Table S5. Summary of annotated transposable elements in the two haplotype-resolved assemblies of GD.

Supplementary Table S6. The telomeres and centromeres identified in the GDT2T_hap1 and GDT2T_hap2 genomes.

Supplementary Table S7. Structural variations detected by genome alignment between the two haplotypes of GD.

Supplementary Table S8. Summary of twelve collected apple genomes.

Supplementary Table S9. Statistics on the number of core, dispensable, and private gene families across seven apple accessions.

Supplementary Table S10. SVs were identified among the eleven apple genomes using GDT2T_hap1 as the reference genome.

Supplementary Table S11. Inversions were identified in wild and cultivated apples.

Supplementary Table S12. Genes and their functional description in the inversion region of chromosome 11.

Supplementary Table S13. The identification and classification of RAGs among the twelve apple genomes.

Supplementary Table S14. Counts of RGAs on different chromosomes in the genome.

Supplementary Table S15. RGAs that have been functionally verified in previous studies.

Supplementary Table S16. The number of NLRs, clustered NLR, and paired NLR genes in each apple genome.

Supplementary Table S17. Expansion of RGAs in cultivated apples.

Supplementary Table S18. Genetic distance between GD and three wild apples.

Supplementary Table S19. Description of genes on GD genome introgression fragments.

Supplementary Table S20. Annotation information of mutations on the introgressed gene *MDgdA02G000320*.

Funding

This study was supported by the Open project of Xinjiang Key Laboratory of Biological Resources and Genetic Engineering (grant 2020D04033 to Xinmin Tian), the Science Fund Program for Distinguished Young Scholars of the National Natural Science Foundation of China (Overseas) to Yongfeng Zhou, and the National Natural Science Foundation of China (grant 32360058 to Xinmin Tian).

Acknowledgments

We are grateful for all members of the Zhou Lab for useful discussions and constructive comments.

Author contributions

X.T and Y.Z. conceived and designed this project; X.T collected plant material. Y.S. collected the genome assemblies; S.C. performed the gene annotation; Y.S., Z.L., and X.Y. performed the data analysis; J.L., Y.W., and X.W. assisted in bioinformatics analyses; Q.L., Z.L., and S.H. provided good guidance for data analysis; Y.S. wrote the manuscript; X.T, Y.Z., H.X. and Z.L revised the manuscript; Y.G. provided images of the apple fruit; Z.C., Y.P., F.Z., H.X., X.C., M.Z., G.Y., and Z.C. checked the grammars; All authors revised and approved the final manuscript.

Data availability

All sequence data along with the assembly and annotation of the GDT2T genome have been deposited in the NCBI Sequence Read Archive under project number PRJNA983958 and PRJNA994290. The assembly and its annotation have been deposited in zenodo: 10.5281/zenodo.10816179 and 10.5281/zenodo.10816304 (https://zenodo.org/record/). The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

Conflict of interest statement. The authors declare that they have no conflict of interest.

Table 1. A comparison of genomic features of GDDH13 and GDT2T_hap1/GDT2T_hap2 reference genomes.

Assembly	GDDH13	GDT2T_hap1	GDT2T_hap2
Length of assembled contigs (bp)	-	663,491,465	663,492,220
Contig N50 (bp)	620,000	37,717,778	34,588,713
Contig BUSCO (%)	-	98.3	98.9
Length of assembled contigs (bp)	-	650,653,859	644693502
Scaffold N50 (bp)	5,500,000	37,717,778	36,267,156
Longest scaffold (bp)	-	56,139,794	54,725,691
Scaffold gaps (>1bp)	3109	9	10
Number of gaps	3109	0	0
Final genome size (bp)	656,833,168	650,775,246	644,751,625
HiFi mapping rate (%)	-	99.5	99.5
Genome BUSCO (%)	96.8	98.7	98.6
GC content (%)	33.55	38.08	38.05
Number of telomeres	-	29/34	29/34
Number of centromeres	-	17	17
Protein-coding genes	42,140	44,375	44,304
Number of transcripts	52,741	63,512	63,437
mRNA BUSCO (%)	94.9	96.6	97.0
Proteins BUSCO (%)	94.9	96.4	97.0
TE (%)	57.3	61.5	61.5

'-', no data in the original publication.

Figure Legends

Figure 1. The haplotype-resolved T2T reference genome of the *Malus domestica* cv. Golden Delicious (GDT2T). **A)** The image of GD. **B)** Anatomy image of GD. The image was digitally extracted for comparison. **C)** Comparative genomics between GDDH13 and two haplotypes of GDT2T. The purple distribution map shows the gene density of GDT2T haplotype 1. Two large inversions (orange color) on chromosomes 1 and 2 indicate the resolution of assembly errors present in GDDH13. The red box on chromosome 10 indicates enhanced completeness of the new assembly. **D)** Hi-C interaction matrix based on the GD diploid assemblies. **E)** Multiple genes are located in the insertion region on chromosome 10 of GDT2T_hap1 as compared to GDDH13. **F)** GO terms of the hemizygous genes.

Figure 2. Pangenome across seven apple species. **A)** Phylogenetic relationships and presence/absence of pan-gene families across seven accessions. Seven accessions: *Malus sieversii* (Msi), *Malus sylvestris* (Msy), *Malus baccata* (Mba), HFTH1, Honeycrisp (HC), Gala, Golden delicious (GD). **B)** Distribution of the number of pan/core gene families for each additional apple genome. **C)** Histogram shows the number of gene families with different frequencies in the seven genomes. The pie chart shows the proportions of different gene family categories. **D)** Violin plot of d_N/d_S distribution in the core, dispensable, and private gene clusters. **E)** The percentage of the three gene clusters with and without protein domain annotations. F) GO analysis of expanded gene families in cultivated (CV) species. **G)** GO analysis of expanded gene families in wild species.

Figure 3. Genome-wide detection of resistance gene analogs (RGAs) across wild and cultivated (CV) species. **A)** Counts of each category in the classification of RGAs based on protein domains. RPW8-X, RPW8-like; CN, CC-NB; TN, TIR-NB; NBS, NB domain only; NL, NB-LRR; TX, TIR-X; TNL, TIR-NB-LRR; RLP, LRR-RLPs; CNL, CC-NB-LRR; RLK, LRR-RLKs; TM-CC, transmembrane-CC (see Methods). **B)** Number of RGAs, clustered NLR, clustered genes, paired NLR and paired genes in wild and cultivated species. The upper and lower edges of the boxes represent the 75%

and 25% quartiles, respectively, while the central line denotes the median. **C)** Diversity distribution of different RGA types across chromosome 1. Different categories of RGAs are displayed with triangles of various colors. **D)** Counts of five duplication events for RGAs of wild and cultivated species. Significant *P*-values between RGAs in wild and cultivated apples: $*P < 0.05$. $**P < 0.01$. All *P*-values obtained by Student's t-tests are indicated. **E)** The proportion of four classifications of structure variations in the genomic region (2 kb) near RGAs. DEL, deletion; DUP, duplication; INV, inversion; INS, insertion. **F)** Comparative heatmap of RGAs families between wild and cultivated species.

Figure 4. Comparative and introgression analyses of RGAs in wild and cultivated apples. **A-B)** The Venn diagram shows the shared and unique RGAs between three wild and one cultivated apple. A: Golden Delicious (GD) with *Malus sieversii* (Msi), *Malus sylvestris* (Msy) and *Malus baccata* (Mba). B: Gala with Msi, Msy, and Mba. **C)** Alignment analysis of genome collinearity among six haplotypes. Different RGA classes are represented by different colored triangle patterns. The red box and arrow indicate the potential introgression of a gene fragment from Mba to GD. **D)** The upper panel shows the genetic distance between the two haplotypes of GD. The middle panel represents the genetic distance between GDT2T_hap1 and the consensus sequences of Msi, Msy, and Mba. The lower panel represents the genetic distance between GDT2T_hap2 and the consensus sequences of Msi, Msy and Mba. Introgressed regions are displayed using gray blocks.

References

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022:**23**(1):1-19. https://doi.org/10.1186/s13059-022-02823-7

Arya P, Kumar G, Acharya V, Singh AK. Genome-wide identification and expression analysis of NBS-encoding genes in *Malus × domestica* and expansion of NBS genes family in Rosaceae. PLoS One. 2014:**9**(9):e107987. https://doi.org/10.1371/journal.pone.0107987

Baurens F, Martin G, Hervouet C, Salmon F, Yohomé D, Ricci S, Rouard M, Habas R, Lemainque, A, Yahiaoui N, et al. Recombination and large structural variations shape interspecific edible bananas genomes. Mol Biol Evol. 2019:36(1):97-111. https://doi.org/10.1093/molbev/msy199

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:**30**(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015:**12**(1):59-60. https://doi.org/10.1038/nmeth.3176

Bus VG, Bassett HC, Bowatte D, Chagné D, Ranatunga CA, Ulluwishewa D, Wiedow C, Gardiner SE. Genome mapping of an apple scab, a powdery mildew and a woolly apple aphid resistance gene from open-pollinated Mildew Immune Selection. Tree Genet Genomes. 2010:**6**:477-487. https://doi.org/10.1007/s11295-009-0265-2

Bus VG, Rikkerink EH, Caffier V, Durel CE, Plummer KM. Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*.Annu Rev Phytopathol. 2011:**49**:391-413. https://doi.org/10.1146/annurev-phyto-072910-095339

Chen J, Liu Y, Liu M, Guo W, Wang Y, He Q, Chen W, Liao Y, Zhang W, Gao Y, et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. Nat Genet. 2023a:55(12):2243-2254.

https://doi.org/10.1038/s41588-023-01571-z

Caffier V, Le Cam B, Al Rifaï M, Bellanger MN, Comby M, Denancé C, Didelot F, Expert P, Kerdraon T, Lemarquand A, et al. Slow erosion of a quantitative apple resistance to *Venturia inaequalis* based on an isolate-specific quantitative trait locus. Infect Genet Evol. 2016:**44**:541-548. https://doi.org/10.1016/j.meegid.2016.07.016

Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, Hu J, Wang K, Wang C, Xin B, et al. A complete telomere-to-telomere assembly of the maize genome. Nat Genet. 2023:**55**(7):1221-1231. https://doi.org/10.1360/tb-2023-0627

Chen P, Li Z, Zhang D, Shen W, Xie Y, Zhang J, Jiang L, Li X, Shen X, Geng D, et al. Insights into the effect of human civilization on *Malus* evolution and domestication. Plant Biotechnol J. 2021:**19**(11):2206-2220. https://doi.org/10.1111/pbi.13648

Chen X, Li S, Zhang D, Han M, Jin X, Zhao C, Wang S, Xing L, Ma J, Ji J. Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. G3: Genes, Genomes, Genetics. 2019:**9**(7):2051-2060. https://doi.org/10.1534/g3.119.400245

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:**18**(2):170-175. https://doi.org/10.1038/s41592-020-01056-5

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012:**6**(2):80-92. https://doi.org/10.4161/fly.19695

Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. Trends Genet. 2014:**30**(2):57-65. https://doi.org/10.1016/j.tig.2013.10.002

Cornille A, Gladieux P, Smulders MJ, Roldán Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, et al. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. Plos Genet. 2012:**8**(5):e1002703.

https://doi.org/10.1371/journal.pgen.1002703

Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, Van de Geest H, Bianco L, Micheletti D, Velasco R, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat Genet. 2017:**49**(7):1099-1106. https://doi.org/10.1038/ng.3886

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021:**10**(2):giab008. https://doi.org/10.1093/gigascience/giab008

Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. Genome Biol. 2021:**22**(1):1-19. https://doi.org/10.1186/s13059-020-02224-8

Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst van Putten HJ, Kruisselbrink JW, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, et al. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res. 2016:**3**:16057. https://doi.org/10.1038/hortres.2016.57

Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, Wang X, Jiao C, Legall N, Mao L, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Nat Commun. 2017:**8**(1):249. https://doi.org/10.1038/s41467-017-00336-7

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017:**356**(6333):92-95. https://doi.org/10.1126/science.aal3327

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016a:**3**(1):99-101. https://doi.org/10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer

provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b:**3**(1):95-98. https://doi.org/10.1016/j.cels.2016.07.002

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:**32**(5):1792-1797. https://doi.org/10.1093/nar/gkh340

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:**20**:1-14. https://doi.org/10.1186/s13059-019-1832-y

Fajkus J, Sýkorová E, Leitch AR. Telomeres in evolution and evolution of telomeres. Chromosome Res. 2005:**13**:469-479. https://doi.org/10.1201/9781498713498-16

Feechan A, Anderson C, Torregrosa L, Jermakow A, Mestre P, Wiedemann Merdinoglu S, Merdinoglu D, Walker AR, Cadle Davidson L, Reisch B, et al. Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. Plant J. 2013:**76**(4):661-674. https://doi.org/10.1111/tpj.12327

Fernandes JB, Wlodzimierz P, Henderson IR. Meiotic recombination within plant centromeres. Curr opin plant biol. 2019:**48**:26-35. https://doi.org/10.1016/j.pbi.2019.02.008

Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. Computational and Structural Biotechnol J. 2020:**18**:9-19. https://doi.org/10.1016/j.csbj.2019.11.002

Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 2022:**38**(10):2922-2926. https://doi.org/10.1093/bioinformatics/btac661

Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019:**20**(1):1-13. https://doi.org/10.1186/s13059-019-1911-0

Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J, Schwab W. Multi-substrate flavonol O-glucosyltransferases from strawberry (*Fragaria*× *ananassa*) achene and receptacle. J Exp Bot. 2008:**59**(10):2611-2625.

https://doi.org/10.1093/jxb/ern117

Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput Sci. 2020:**6**:e251. https://doi.org/10.7287/peerj.preprints.27928

Heitz T, Widemann E, Lugan R, Miesch L, Ullmann P, Désaubry L, Holder E, Grausem B, Kandel S, Miesch M, et al. Cytochromes P450 *CYP94C1* and *CYP94B3* catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. J Biol Chem. 2012:**287**(9):6296-6306. https://doi.org/10.1074/jbc.m111.316364

Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics. 2019:**35**(17):2907-2915. https://doi.org/10.1093/bioinformatics/btz041

Huang H, Liu X, Arshad R, Wang X, Li WM, Zhou Y, Ge XJ. Telomere-to-telomere haplotype-resolved reference genome reveals subgenome divergence and disease resistance in triploid Cavendish banana. Hortic Res. 2023:**10**(9):uhad153. https://doi.org/10.1093/hr/uhad153

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017:**8**(1):14061. https://doi.org/10.1038/ncomms14061

Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, Liu Y, Liu B, Wang Y. Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. 2020:**21**(1):1-24. https://doi.org/10.1186/s13059-020-02107-y

Karlström A, Gómez-Cortecero A, Nellist CF, Ordidge M, Dunwell JM, Harrison RJ. Identification of novel genetic regions associated with resistance to European canker in apple. Bmc Plant Biol. 2022:**22**(1):452. https://doi.org/10.1101/2021.12.20.473552

KB N. GeneDoc: analysis and visualization of genetic variation. EMBnet news. 1997:**4**:1-4.

Khan A, Carey SB, Gomez AS, Zhang H, Hargarten H, Hale H, Harkess A, Honaas L. A phased, chromosome-scale genome of 'Honeycrisp' apple (*Malus domestica*).

Gigabyte. 2022. https://doi.org/10.46471/gigabyte.69

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019:**37**(8):907-915. https://doi.org/10.1038/s41587-019-0201-4

Lee S, Seo J, Rodriguez-Lanetty M, Lee DH. Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. Mol Genet Genomics. 2003:**269**(1):101-108. https://doi.org/10.1007/s00438-003-0816-4

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:**34**(18):3094-3100. https://doi.org/10.1093/bioinformatics/bty191

Li H, Wang S, Chai S, Yang Z, Zhang Q, Xin H, Xu Y, Lin S, Chen X, Yao Z, et al. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. Nat Commun. 2022a:**13**(1):682. https://doi.org/10.1038/s41467-022-28362-0

Li N, He Q, Wang J, Wang B, Zhao J, Huang S, Yang T, Tang Y, Yang S, Aisimutuola P, et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. Nat Genet. 2023:**55**(5):852-860. https://doi.org/10.1038/s41588-023-01340-y

Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC genomics. 2016:**17**(1):1-10. https://doi.org/10.1186/s12864-016-3197-x

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006:**22**(13):1658-1659. https://doi.org/10.1093/bioinformatics/btl158

Li Y, Cao K, Li N, Zhu G, Fang W, Chen C, Wang X, Guo J, Wang Q, Ding T, et al. Genomic analyses provide insights into peach local adaptation and responses to climate change. Genome Res. 2021:**31**(4):592-606. https://doi.org/10.1101/gr.261032.120

Li Z, Wang L, He J, Li X, Hou N, Guo J, Niu C, Li C, Liu S, Xu J, et al. Chromosome-

scale reference genome provides insights into the genetic origin and grafting-mediated stress tolerance of *Malus prunifolia*. Plant Biotechnol J. 2022:**20**(6):1015-1017. https://doi.org/10.1111/pbi.13817

Liao L, Zhang W, Zhang B, Fang T, Wang XF, Cai Y, Ogutu C, Gao L, Chen G, Nie X, et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. Mol Plant. 2021:**14**(9):1454-1471. https://doi.org/10.1016/j.molp.2021.05.018

Liao W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. A draft human pangenome reference. Nature. 2023:**617**(7960):312-324. https://doi.org/10.1530/ey.20.12.1

Lin Q, Chen J, Liu X, Wang B, Zhao Y, Liao L, Allan AC, Sun C, Duan Y, Li X, et al. A metabolic perspective of selection for fruit quality related to apple domestication and improvement. Genome Biol. 2023:**24**(1):1-22. https://doi.org/10.1186/s13059-023-02945-6

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, et al. Pan-genome of wild and cultivated soybeans. Cell. 2020:**182**(1):162-176. https://doi.org/10.1016/j.cell.2020.05.023

Liu Z, Ren Z, Yan L, Li F. DeepLRR: An Online Webserver for Leucine-Rich-Repeat Containing Protein Characterization Based on Deep Learning. Plants. 2022:**11**(1):136. https://doi.org/10.3390/plants11010136

Long Q, Cao S, Huang G, Wang X, Liu Z, Liu W, Wang Y, Xiao H, Peng Y, Zhou Y. Population comparative genomics discovers gene gain and loss during grapevine domestication. plant physiol. 2024:**00**:1-13. https://doi.org/10.1093/plphys/kiae039

Luo F, Evans K, Norelli JL, Zhang Z, Peace C. Prospects for achieving durable disease resistance with elite fruit quality in apple breeding. Tree Genet Genomes. 2020a:**16**:1-14. https://doi.org/10.1007/s11295-020-1414-x

Luo F, Norelli JL, Howard NP, Wisniewski M, Flachowsky H, Hanke MV, Peace C. Introgressing blue mold resistance into elite apple germplasm by rapid cycle breeding and foreground and background DNA-informed selection. Tree Genet Genomes. 2020b:**16**:1-15. https://doi.org/10.1007/s11295-020-1419-5

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 2017:**33**(4):574-576. https://doi.org/10.1093/bioinformatics/btw663

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018:**14**(1):e1005944. https://doi.org/10.1371/journal.pcbi.1005944

McDonald BA, Stukenbrock EH. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. Phil Trans R Soc B 2016:**371**(1709):20160026. https://doi.org/10.1098/rstb.2016.0026

Morales-Cruz A, Aguirre-Liguori J, Massonnet M, Minio A, Zaccheo M, Cochetel N, Walker A, Riaz S, Zhou Y, Cantu D, et al. Multigenic resistance to *Xylella fastidiosa* in wild grapes (*Vitis* sps.) and its implications within a changing climate. Commun Biol. 2023:**6**(1):580. https://doi.org/10.1038/s42003-023-04938-4

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:**32**(1):268-274. https://doi.org/10.1093/molbev/msu300

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. The complete sequence of a human genome. Science. 2022:**376**(6588):44-53. https://doi.org/10.1126/science.abj6987

O'Rourke D. Economic importance of the world apple industry. The Apple Genome. 2021:1-18. https://doi.org/10.1007/978-3-030-74682-7_1

Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019:**20**(1):1-18. https://doi.org/10.1186/s13059-019-1905-y

Peil A, Garcia-Libreros T, Richter K, Trognitz F, Trognitz B, Hanke MV, Flachowsky H. Strong evidence for a fire blight resistance gene of *Malus robusta* located on linkage group 3. Plant Breeding. 2007:**126**(5):470-475. https://doi.org/10.1111/j.1439-0523.2007.01408.x

Perazzolli M, Malacarne G, Baldo A, Righetti L, Bailey A, Fontana P, Velasco R, Malnoy M. Characterization of resistance gene analogues (RGAs) in apple (*Malus* × *domestica* Borkh.) and their evolutionary history of the Rosaceae family. PloS one. 2014:**9**(2):e83844. https://doi.org/10.1371/journal.pone.0083844

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 2015:**33**(3):290-295. https://doi.org/10.1038/nbt.3122

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007:**81**(3):559-575. https://doi.org/10.1086/519795

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:**26**(6):841-842. https://doi.org/10.1093/bioinformatics/btq033

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:**11**(1):1432. https://doi.org/10.1038/s41467-020-14998-3

Schnable PS, Springer NM. Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol. 2013:**64**:71-88. https://doi.org/10.1146/annurev-arplant-042110-103827

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018:**15**(6):461-468. https://doi.org/10.1038/s41592-018-0001-7

Shang L, He W, Wang T, Yang Y, Xu Q, Zhao X, Yang L, Zhang H, Li X, Lv Y, et al. A complete assembly of the rice Nipponbare reference genome. Mol Plant. 2023:**16**(8):1232-1236. https://doi.org/10.1016/j.molp.2023.08.003

Shang L, Li X, He H, Yuan Q, Song Y, Wei Z, Lin H, Hu M, Zhao F, Zhang C, et al. A super pan-genomic landscape of rice. Cell Res. 2022:**32**(10):878-896. https://doi.org/10.1038/s41422-022-00685-z

Sherman RM, Salzberg SL. Pan-genomics in the human genome era. Nat Rev Genet. 2020:**21**(4):243-254. https://doi.org/10.1038/s41576-020-0210-7

Shi X, Cao S, Wang X, Huang S, Wang Y, Liu Z, Liu W, Leng X, Peng Y, Wang N, et al. The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. Hortic Res. 2023:**10**(05):uhad061. https://doi.org/10.1093/hr/uhad061

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021:**37**(12):1639-1643. https://doi.org/10.1093/bioinformatics/btaa1016

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:**31**(19):3210-3212. https://doi.org/10.1093/bioinformatics/btv351

Singh J, Sun M, Cannon SB, Wu J, Khan A. An accumulation of genetic variation and selection across the disease-related genes during apple domestication. Tree Genet Genomes. 2021:**17**:1-11. https://doi.org/10.1007/s11295-021-01510-1

Song JM, Xie WZ, Wang S, Guo YX, Koo DH, Kudrna D, Gong C, Huang Y, Feng JW, Zhang W, et al. Two gap-free reference genomes and a global view of the centromere architecture in rice. Mol Plant. 2021:**14**(10):1757-1767. https://doi.org/10.1016/j.molp.2021.06.018

Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, Baggs E, Witek AI, Yadav I, Krasileva KV, et al. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. Plant Physiol. 2020:**183**(2):468-482. https://doi.org/10.1104/pp.19.01273

Sun M, Yao C, Shu Q, He Y, Chen G, Yang G, Xu S, Liu Y, Xue Z, Wu J. Telomere-to-telomere pear (*Pyrus pyrifolia*) reference genome reveals segmental and whole genome duplication driving genome evolution. Hortic Res. 2023:**10**(11):uhad201. https://doi.org/10.1093/hr/uhad201

Sun M, Zhang M, Singh J, Song B, Tang Z, Liu Y, Wang R, Qin M, Li J, Khan A, et al. Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears. BMC genomics. 2020a:**21**:1-16.

https://doi.org/10.1186/s12864-020-07226-1

Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, Khan A, Ban S, Xu K, Cheng L, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. Nat Genet. 2020b:**52**(12):1423-1432. https://doi.org/10.1038/s41588-020-00723-9

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021:**38**(7):3022-3027. https://doi.org/10.1093/molbev/msab120

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, et al. Genome evolution and diversity of wild and cultivated potatoes. Nature. 2022:**606**(7914):535-541. https://doi.org/10.1038/s41586-022-04822-x

Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J, et al. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 2015:**16**:1-15. https://doi.org/10.1186/s13059-014-0573-1

Van de Weyer A, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JD, Dangl JL, Weigel D, Bemm F, et al. A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. Cell. 2019:**178**(5):1260-1272. https://doi.org/10.1016/j.cell.2019.07.038

Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 5Gs for crop genetic improvement. Curr opin plant biol. 2020:**56**:190-196. https://doi.org/10.1016/j.pbi.2019.12.004

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat Genet. 2010:**42**(10):833-839. https://doi.org/10.1038/ng.654

Wang N, Cao S, Liu Z, Xiao H, Hu J, Xu X, Chen P, Ma Z, Ye J, Chai L, et al. Genomic conservation of crop wild relatives: A case study of citrus. Plos Genet. 2023:**19**: e1010811. https://doi.org/10.1371/journal.pgen.1010811

Wang T, Duan S, Xu C, Wang Y, Zhang X, Xu X, Chen L, Han Z, Wu T, et al.

Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits. Nat Commun. 2023:**14**(1):7377. https://doi.org/10.1038/s41467-023-43270-7

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012:**40**(7):e49-e49. https://doi.org/10.1093/nar/gkr1293

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The innovation. 2021:**2**(3). https://doi.org/10.1016/j.xinn.2021.100141

Xiao H, Liu Z, Wang N, Long Q, Cao S, Huang G, Liu W, Peng Y, Riaz S, Walker AM, et al. Adaptive and maladaptive introgression in grapevine domestication. Proc Nat Acad Sci. 2023:**120**(24):e2222041120. https://doi.org/10.1073/pnas.2222041120

Xie D, Xu Y, Wang J, Liu W, Zhou Q, Luo S, Huang W, He X, Li Q, Peng Q, et al. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun. 2019:**10**(1):5158. https://doi.org/10.1038/s41467-019-13185-3

Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen JQ, Hartl DL, Tian D. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Nat Acad Sci. 2013:**110**(46):18572-18577. https://doi.org/10.1073/pnas.1318211110

Yang T, Liu R, Luo Y, Hu S, Wang D, Wang C, Pandey MK, Ge S, Xu Q, Li N, et al. Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. Nat Genet. 2022:**54**(10):1553-1563. https://doi.org/10.1038/s41588-022-01172-2

Yue J, Chen Q, Wang Y, Zhang L, Ye C, Wang X, Cao S, Lin Y, Huang W, Xian H, et al. Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit *Actinidia chinensis*. Hortic Res. 2023:**10**(2):uhac264. https://doi.org/10.1093/hr/uhac264

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al.

A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun. 2019a:**10**(1):1494. https://doi.org/10.1038/s41467-019-09518-x

Zhang Q, Wang Y, Wei H, Fan W, Xu C, Li T. CCR-NB-LRR proteins *MdRNL2* and *MdRNL6* interact physically to confer broad-spectrum fungal resistance in apple (*Malus* × *domestica*). Plant J. 2021:**108**(5):1522-1538. https://doi.org/10.1111/tpj.15526

Zhang Y, Guo M, Shen J, Song X, Dong S, Wen Y, Yuan X, Guo P. Comparative genomics analysis in grass species reveals two distinct evolutionary strategies adopted by R genes. Sci Rep. 2019b:**9**(1):10735. https://doi.org/10.1038/s41598-019-47121-8

Zheng W, Shen F, Wang W, Wu B, Wang X, Xiao C, Tian Z, Yang X, Yang J, Wang Y, et al. Quantitative trait loci-based genomics-assisted prediction for the degree of apple fruit cover color. The Plant Genome. 2020:**13**(3):e20047. https://doi.org/10.1002/tpg2.20047

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, Cantu D, Gaut BS. The population genetics of structural variants in grapevine domestication. Nature plants. 2019:**5**(9):965-979. https://doi.org/10.1038/s41477-019-0507-8

Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y, Wu K, et al. Graph pangenome captures missing heritability and empowers tomato breeding. Nature. 2022:**606**(7914):527-534. https://doi.org/10.1038/s41586-022-04808-9

Figure 1

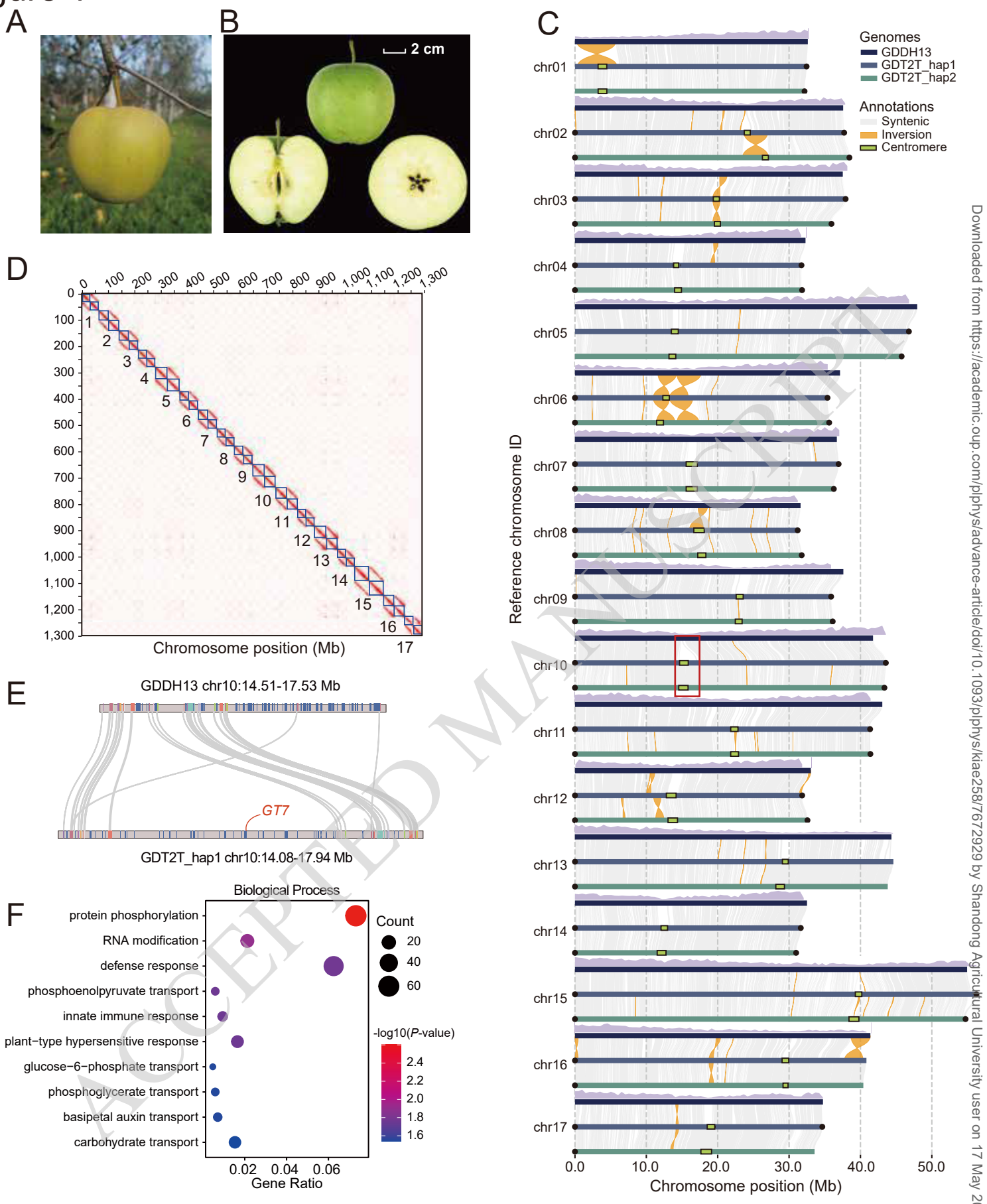

Figure 1. The haplotype-resolved T2T reference genome of the 'Golden Delicious' (GDT2T). **A)** The image of GD. **B)** Anatomy image of GD. The image were digitally extracted for comparison. **C)** Comparative genomics between GDDH13 and two haplotypes of GDT2T. The purple distribution map shows the gene density of GDT2T haplotype 1. Two large inversions on chromosomes (chr) 1 and 2 indicate the resolution of assembly errors present in GDDH13. The red box on chromosome 10 indicates enhanced completeness of the new assembly. **D)** Hi-C interaction matrix based on the GD diploid assemblies. **E)** Multiple genes are located in the insertion region on chromosome 10 of GDT2T_hap1 compared to GDDH13. **F)** GO terms of the hemizygous genes.

Figure 2

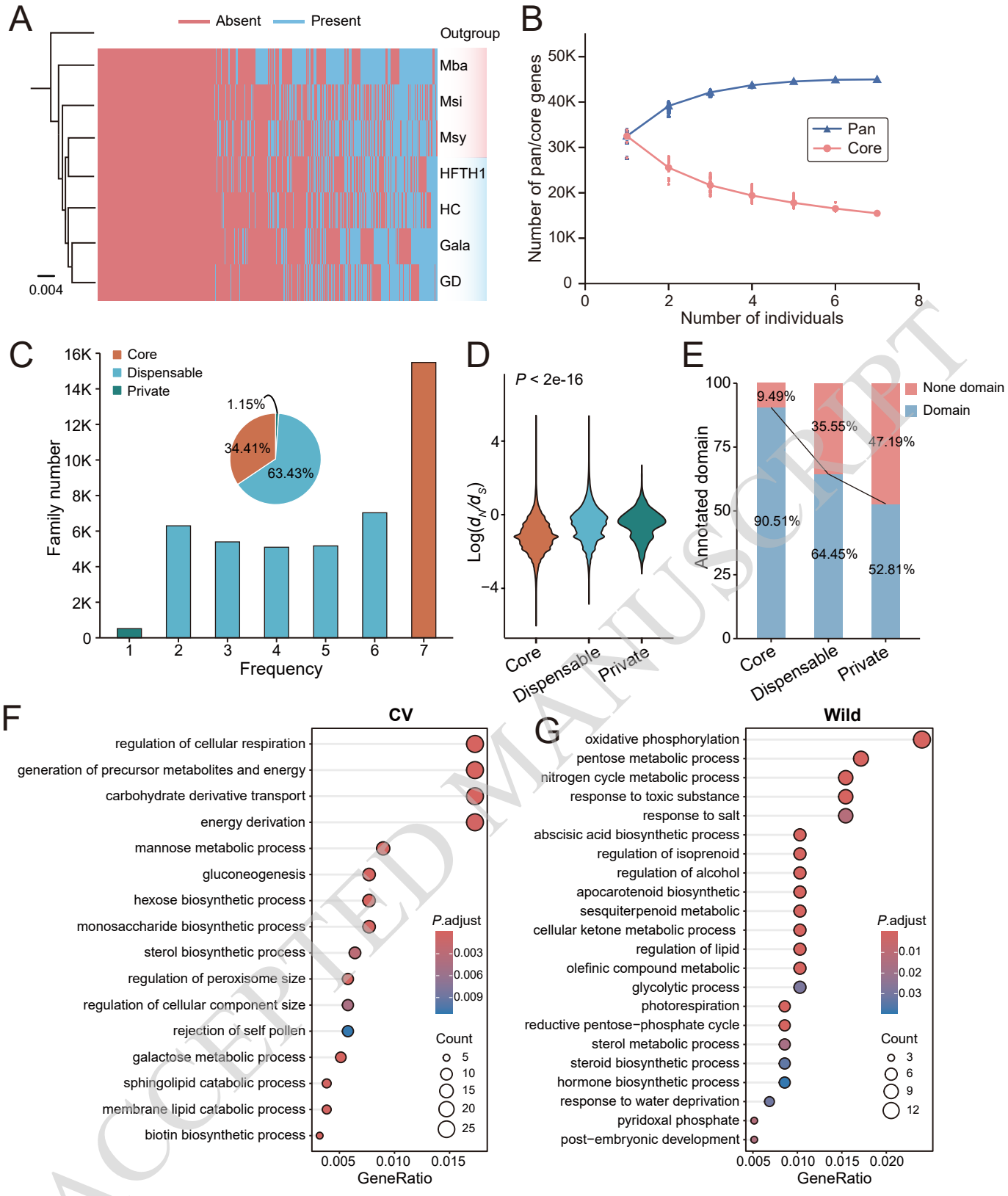

Figure 2. Pangenome across seven apple species. **A)** Phylogenetic relationships and presence/absence of pan-gene families across seven accessions. Seven accessions: *Malus sieversii* (Msi), *Malus sylvestris* (Msy), *Malus baccata* (Mba), HFTH1, Honeycrisp (HC), Gala, Golden delicious (GD). **B)** Distribution of the number of pan/core gene families for each additional apple genome. **C)** Histogram shows the number of gene families with different frequencies in the seven genomes. The pie chart shows the proportions of different gene family categories. **D)** Violin plot of d_N/d_S distribution in the core, dispensable, and private gene clusters. **E)** The percentage of the three gene clusters with and without protein domain annotations. **F)** GO analysis of expanded gene families in cultivated (CV) species. **G)** GO analysis of expanded gene families in wild species.

Figure 3

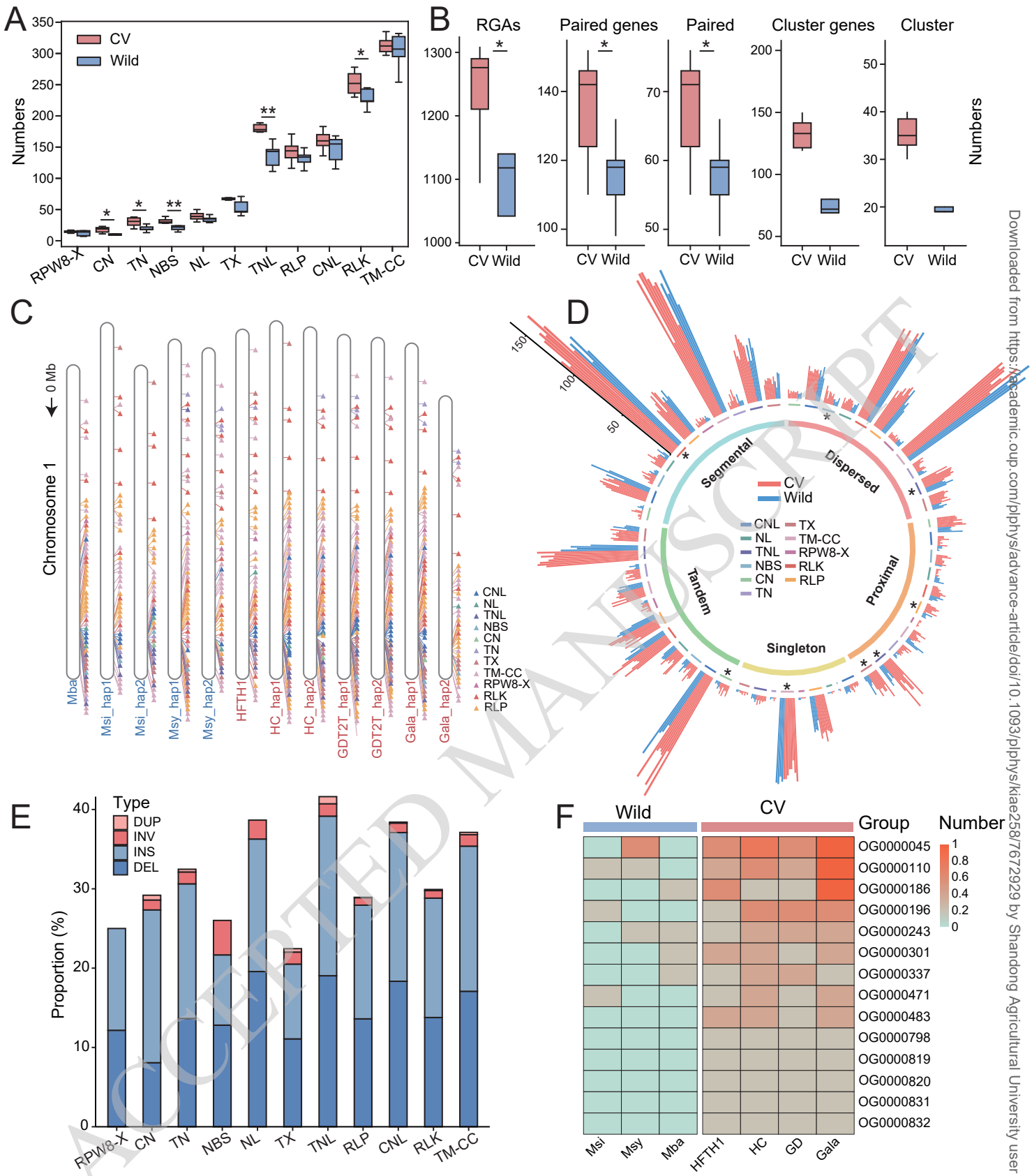

Figure 3. Genome-wide detection of resistance gene analogs (RGAs) across wild and cultivated (CV) species. **A)** Counts of each category in the classification of RGAs based on protein domains. RPW8-X, RPW8-like; CN, CC-NB; TN, TIR-NB; NBS, NB domain only; NL, NB-LRR; TX, TIR-X; TNL, TIR-NB-LRR; RLP, LRR-RLPs; CNL, CC-NB-LRR; RLK, LRR-RLKs; TM-CC, transmembrane-CC (see Methods). **B)** Number of RGAs, clustered NLR, clustered genes, paired NLR and paired genes in wild and cultivated species. The upper and lower edges of the boxes represent the 75% and 25% quartiles, respectively, while the central line denotes the median. **C)** Diversity distribution of different RGAs types across chromosome 1. Different categories of RGAs are displayed with triangles of various colors. **D)** Counts of five duplication events for RGAs of wild and cultivated species. Significant *P*-values between RGAs in wild and cultivated apples: $*P < 0.05$. $**P < 0.01$. All *p*-values obtained by Student's t-tests are indicated. **E)** The proportion of four classifications of structure variations in the genomic region (2 kb) near RGAs. DEL, deletion; DUP, duplication; INV, inversion; INS, insertion. **F)** Comparative heatmap of RGAs families between wild and cultivated species.

Figure 4

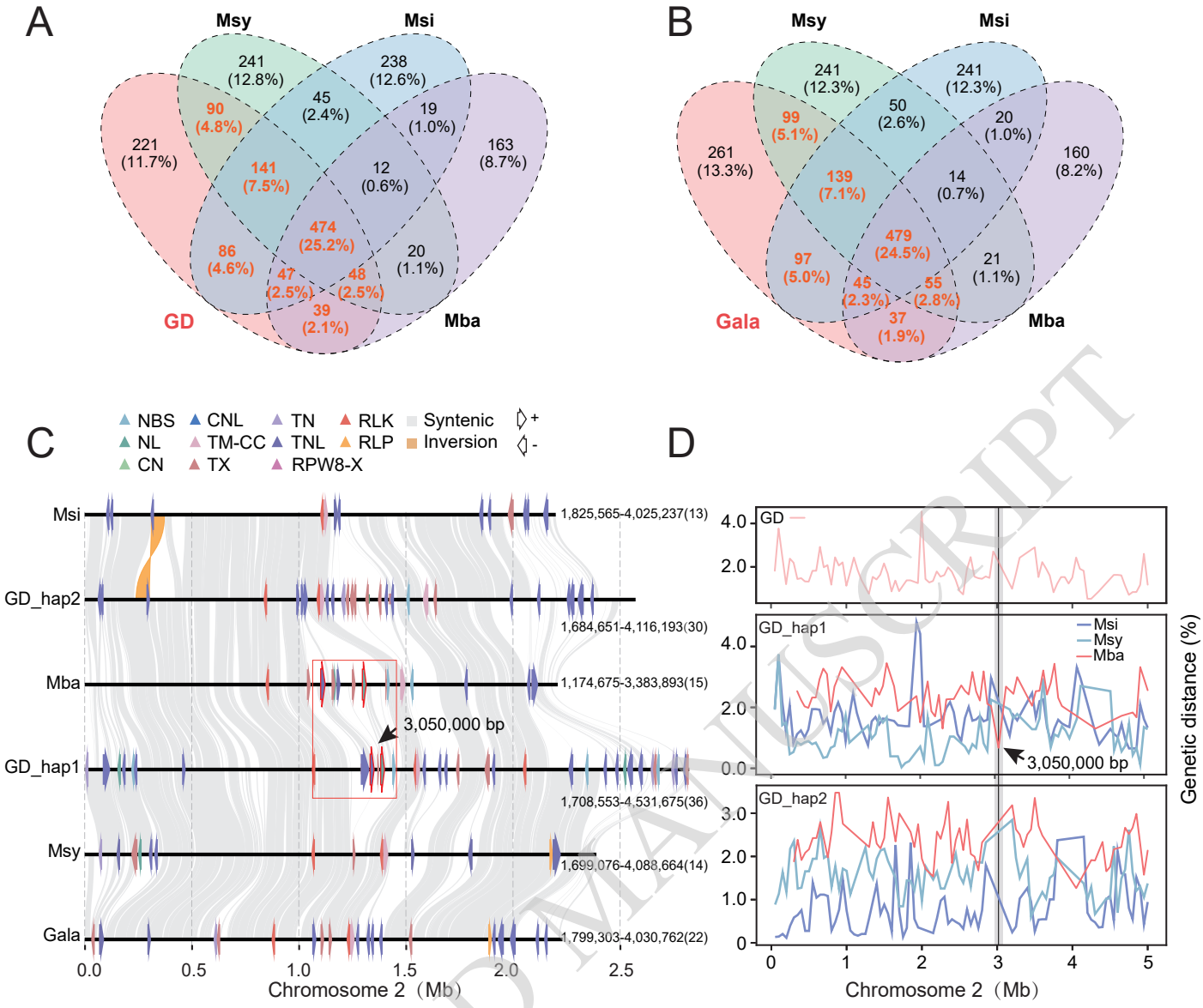

Figure 4. Comparative and introgression analyses of RGAs in wild and cultivated apples. A-B) The Venn diagram shows the shared and unique RGAs between three wild and one cultivated apple. A: Golden Delicious (GD) with *Malus sieversii* (Msi), *Malus sylvestris* (Msy), and *Malus baccata* (Mba). B: Gala with Msi, Msy, and Mba. **C)** Alignment analysis of genome collinearity among six haplotypes. Different RGA classes are represented by different colored triangle patterns. The red box and arrow indicate potential introgression of a gene fragment from Mba to GD. **D)** The upper panel shows the genetic distance between the two haplotypes of GD. The middle panel represent the genetic distance between GDT2T_hap1 and the consensus sequences of Msi, Msy, and Mba. The lower panel represents the genetic distance between GDT2T_hap2 and the consensus sequences of Msi, Msy, and Mba. Introgressed regions are displayed using gray blocks.

Parsed Citations

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022:23(1):1-19. https://doi.org/10.1186/s13059-022-02823-7
Google Scholar: Author Only Title Only Author and Title

Arya P, Kumar G, Acharya V, Singh AK. Genome-wide identification and expression analysis of NBS-encoding genes in Malus × domestica and expansion of NBS genes family in Rosaceae. PLoS One. 2014:9(9):e107987. https://doi.org/10.1371/journal.pone.0107987
Google Scholar: Author Only Title Only Author and Title

Baurens F, Martin G, Hervouet C, Salmon F, Yohomé D, Ricci S, Rouard M, Habas R, Lemainque, A, Yahiaoui N, et al. Recombination and large structural variations shape interspecific edible bananas genomes. Mol Biol Evol. 2019:36(1):97-111. https://doi.org/10.1093/molbev/msy199
Google Scholar: Author Only Title Only Author and Title

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:30(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170
Google Scholar: Author Only Title Only Author and Title

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015:12(1):59-60. https://doi.org/10.1038/nmeth.3176
Google Scholar: Author Only Title Only Author and Title

Bus VG, Bassett HC, Bowatte D, Chagné D, Ranatunga CA, Ulluwishewa D, Wiedow C, Gardiner SE. Genome mapping of an apple scab, a powdery mildew and a woolly apple aphid resistance gene from open-pollinated Mildew Immune Selection. Tree Genet Genomes. 2010:6:477-487. https://doi.org/10.1007/s11295-009-0265-2
Google Scholar: Author Only Title Only Author and Title

Bus VG, Rikkerink EH, Caffier V, Durel CE, Plummer KM. Revision of the nomenclature of the differential host-pathogen interactions of Venturia inaequalis and Malus.Annu Rev Phytopathol. 2011:49:391-413. https://doi.org/10.1146/annurev-phyto-072910-095339
Google Scholar: Author Only Title Only Author and Title

Chen J, Liu Y, Liu M, Guo W, Wang Y, He Q, Chen W, Liao Y, Zhang W, Gao Y, et al. Pangenome analysis reveals genomic variations associated with domestication traits in broomcorn millet. Nat Genet. 2023a:55(12):2243-2254. https://doi.org/10.1038/s41588-023-01571-z
Google Scholar: Author Only Title Only Author and Title

Caffier V, Le Cam B, Al Rifaï M, Bellanger MN, Comby M, Denancé C, Didelot F, Expert P, Kerdraon T, Lemarquand A, et al. Slow erosion of a quantitative apple resistance to Venturia inaequalis based on an isolate-specific quantitative trait locus. Infect Genet Evol. 2016:44:541-548. https://doi.org/10.1016/j.meegid.2016.07.016
Google Scholar: Author Only Title Only Author and Title

Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, Hu J, Wang K, Wang C, Xin B, et al. A complete telomere-to-telomere assembly of the maize genome. Nat Genet. 2023:55(7):1221-1231. https://doi.org/10.1360/tb-2023-0627
Google Scholar: Author Only Title Only Author and Title

Chen P, Li Z, Zhang D, Shen W, Xie Y, Zhang J, Jiang L, Li X, Shen X, Geng D, et al. Insights into the effect of human civilization on Malus evolution and domestication. Plant Biotechnol J. 2021:19(11):2206-2220. https://doi.org/10.1111/pbi.13648
Google Scholar: Author Only Title Only Author and Title

Chen X, Li S, Zhang D, Han M, Jin X, Zhao C, Wang S, Xing L, Ma J, Ji J. Sequencing of a wild apple (Malus baccata) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. G3: Genes, Genomes, Genetics. 2019:9(7):2051-2060. https://doi.org/10.1534/g3.119.400245
Google Scholar: Author Only Title Only Author and Title

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18(2):170-175. https://doi.org/10.1038/s41592-020-01056-5
Google Scholar: Author Only Title Only Author and Title

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012:6(2):80-92. https://doi.org/10.4161/fly.19695
Google Scholar: Author Only Title Only Author and Title

Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. Trends Genet. 2014:30(2):57-65. https://doi.org/10.1016/j.tig.2013.10.002
Google Scholar: Author Only Title Only Author and Title

Cornille A, Gladieux P, Smulders MJ, Roldán Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, et al. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. Plos Genet. 2012:8(5):e1002703. https://doi.org/10.1371/journal.pgen.1002703
Google Scholar: Author Only Title Only Author and Title

Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, Van de Geest H, Bianco L, Micheletti D, Velasco R, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat Genet. 2017:49(7):1099-1106. https://doi.org/10.1038/ng.3886
Google Scholar: Author Only Title Only Author and Title

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008
Google Scholar: Author Only Title Only Author and Title

Della Coletta R, Qiu Y, Ou S, Hufford MB, Hirsch CN. How the pan-genome is changing crop genomics and improvement. Genome Biol. 2021:22(1):1-19. https://doi.org/10.1186/s13059-020-02224-8
Google Scholar: Author Only Title Only Author and Title

Di Pierro EA, Gianfranceschi L, Di Guardo M, Koehorst van Putten HJ, Kruisselbrink JW, Longhi S, Troggio M, Bianco L, Muranty H, Pagliarani G, et al. A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. Hortic Res. 2016:3:16057. https://doi.org/10.1038/hortres.2016.57
Google Scholar: Author Only Title Only Author and Title

Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, Wang X, Jiao C, Legall N, Mao L, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Nat Commun. 2017:8(1):249. https://doi.org/10.1038/s41467-017-00336-7
Google Scholar: Author Only Title Only Author and Title

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017:356(6333):92-95. https://doi.org/10.1126/science.aal3327
Google Scholar: Author Only Title Only Author and Title

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016a:3(1):99-101. https://doi.org/10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b:3(1):95-98. https://doi.org/10.1016/j.cels.2016.07.002

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004:32(5):1792-1797. https://doi.org/10.1093/nar/gkh340
Google Scholar: Author Only Title Only Author and Title

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:1-14. https://doi.org/10.1186/s13059-019-1832-y
Google Scholar: Author Only Title Only Author and Title

Fajkus J, Sýkorová E, Leitch AR. Telomeres in evolution and evolution of telomeres. Chromosome Res. 2005:13:469-479. https://doi.org/10.1201/9781498713498-16
Google Scholar: Author Only Title Only Author and Title

Feechan A, Anderson C, Torregrosa L, Jermakow A, Mestre P, Wiedemann Merdinoglu S, Merdinoglu D, Walker AR, Cadle Davidson L, Reisch B, et al. Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species Muscadinia rotundifolia identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. Plant J. 2013:76(4):661-674. https://doi.org/10.1111/tpj.12327
Google Scholar: Author Only Title Only Author and Title

Fernandes JB, Wlodzimierz P, Henderson IR. Meiotic recombination within plant centromeres. Curr opin plant biol. 2019:48:26-35. https://doi.org/10.1016/j.pbi.2019.02.008
Google Scholar: Author Only Title Only Author and Title

Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. Computational and Structural Biotechnol J. 2020:18:9-19. https://doi.org/10.1016/j.csbj.2019.11.002
Google Scholar: Author Only Title Only Author and Title

Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 2022:38(10):2922-2926. https://doi.org/10.1093/bioinformatics/btac661
Google Scholar: Author Only Title Only Author and Title

Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-

genome assemblies. Genome Biol. 2019:20(1):1-13. https://doi.org/10.1186/s13059-019-1911-0
Google Scholar: Author Only Title Only Author and Title

Griesser M, Vitzthum F, Fink B, Bellido ML, Raasch C, Munoz-Blanco J, Schwab W. Multi-substrate flavonol O-glucosyltransferases from strawberry (Fragaria× ananassa) achene and receptacle. J Exp Bot. 2008:59(10):2611-2625. https://doi.org/10.1093/jxb/ern117
Google Scholar: Author Only Title Only Author and Title

Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput Sci. 2020:6:e251. https://doi.org/10.7287/peerj.preprints.27928
Google Scholar: Author Only Title Only Author and Title

Heitz T, Widemann E, Lugan R, Miesch L, Ullmann P, Désaubry L, Holder E, Grausem B, Kandel S, Miesch M, et al. Cytochromes P450 CYP94C1 and CYP94B3 catalyze two successive oxidation steps of plant hormone jasmonoyl-isoleucine for catabolic turnover. J Biol Chem. 2012:287(9):6296-6306. https://doi.org/10.1074/jbc.m111.316364
Google Scholar: Author Only Title Only Author and Title

Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics. 2019:35(17):2907-2915. https://doi.org/10.1093/bioinformatics/btz041
Google Scholar: Author Only Title Only Author and Title

Huang H, Liu X, Arshad R, Wang X, Li WM, Zhou Y, Ge XJ. Telomere-to-telomere haplotype-resolved reference genome reveals subgenome divergence and disease resistance in triploid Cavendish banana. Hortic Res. 2023:10(9):uhad153. https://doi.org/10.1093/hr/uhad153
Google Scholar: Author Only Title Only Author and Title

Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017:8(1):14061. https://doi.org/10.1038/ncomms14061
Google Scholar: Author Only Title Only Author and Title

Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, Liu Y, Liu B, Wang Y. Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. 2020:21(1):1-24. https://doi.org/10.1186/s13059-020-02107-y
Google Scholar: Author Only Title Only Author and Title

Karlström A, Gómez-Cortecero A, Nellist CF, Ordidge M, Dunwell JM, Harrison RJ. Identification of novel genetic regions associated with resistance to European canker in apple. Bmc Plant Biol. 2022:22(1):452. https://doi.org/10.1101/2021.12.20.473552
Google Scholar: Author Only Title Only Author and Title

KB N. GeneDoc: analysis and visualization of genetic variation. EMBnet news. 1997:4:1-4.
Google Scholar: Author Only Title Only Author and Title

Khan A, Carey SB, Gomez AS, Zhang H, Hargarten H, Hale H, Harkess A, Honaas L. A phased, chromosome-scale genome of 'Honeycrisp' apple (Malus domestica). Gigabyte. 2022. https://doi.org/10.46471/gigabyte.69

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019:37(8):907-915. https://doi.org/10.1038/s41587-019-0201-4
Google Scholar: Author Only Title Only Author and Title

Lee S, Seo J, Rodriguez-Lanetty M, Lee DH. Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. Mol Genet Genomics. 2003:269(1):101-108. https://doi.org/10.1007/s00438-003-0816-4
Google Scholar: Author Only Title Only Author and Title

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094-3100. https://doi.org/10.1093/bioinformatics/bty191
Google Scholar: Author Only Title Only Author and Title

Li H, Wang S, Chai S, Yang Z, Zhang Q, Xin H, Xu Y, Lin S, Chen X, Yao Z, et al. Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. Nat Commun. 2022a:13(1):682. https://doi.org/10.1038/s41467-022-28362-0

Li N, He Q, Wang J, Wang B, Zhao J, Huang S, Yang T, Tang Y, Yang S, Aisimutuola P, et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. Nat Genet. 2023:55(5):852-860. https://doi.org/10.1038/s41588-023-01340-y
Google Scholar: Author Only Title Only Author and Title

Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC genomics. 2016:17(1):1-10. https://doi.org/10.1186/s12864-016-3197-x
Google Scholar: Author Only Title Only Author and Title

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006:22(13):1658-1659. https://doi.org/10.1093/bioinformatics/btl158
Google Scholar: Author Only Title Only Author and Title

Li Y, Cao K, Li N, Zhu G, Fang W, Chen C, Wang X, Guo J, Wang Q, Ding T, et al. Genomic analyses provide insights into peach local adaptation and responses to climate change. Genome Res. 2021:31(4):592-606. https://doi.org/10.1101/gr.261032.120
Google Scholar: Author Only Title Only Author and Title

Li Z, Wang L, He J, Li X, Hou N, Guo J, Niu C, Li C, Liu S, Xu J, et al. Chromosome-scale reference genome provides insights into the genetic origin and grafting-mediated stress tolerance of Malus prunifolia. Plant Biotechnol J. 2022:20(6):1015-1017. https://doi.org/10.1111/pbi.13817
Google Scholar: Author Only Title Only Author and Title

Liao L, Zhang W, Zhang B, Fang T, Wang XF, Cai Y, Ogutu C, Gao L, Chen G, Nie X, et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. Mol Plant. 2021:14(9):1454-1471. https://doi.org/10.1016/j.molp.2021.05.018
Google Scholar: Author Only Title Only Author and Title

Liao W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. A draft human pangenome reference. Nature. 2023:617(7960):312-324. https://doi.org/10.1530/ey.20.12.1
Google Scholar: Author Only Title Only Author and Title

Lin Q, Chen J, Liu X, Wang B, Zhao Y, Liao L, Allan AC, Sun C, Duan Y, Li X, et al. A metabolic perspective of selection for fruit quality related to apple domestication and improvement. Genome Biol. 2023:24(1):1-22. https://doi.org/10.1186/s13059-023-02945-6
Google Scholar: Author Only Title Only Author and Title

Liu Y, Du H, Li P, Shen Y, Peng H, Liu S, Zhou GA, Zhang H, Liu Z, Shi M, et al. Pan-genome of wild and cultivated soybeans. Cell. 2020:182(1):162-176. https://doi.org/10.1016/j.cell.2020.05.023
Google Scholar: Author Only Title Only Author and Title

Liu Z, Ren Z, Yan L, Li F. DeepLRR: An Online Webserver for Leucine-Rich-Repeat Containing Protein Characterization Based on Deep Learning. Plants. 2022:11(1):136. https://doi.org/10.3390/plants11010136
Google Scholar: Author Only Title Only Author and Title

Long Q, Cao S, Huang G, Wang X, Liu Z, Liu W, Wang Y, Xiao H, Peng Y, Zhou Y. Population comparative genomics discovers gene gain and loss during grapevine domestication. plant physiol. 2024:00:1-13. https://doi.org/10.1093/plphys/kiae039
Google Scholar: Author Only Title Only Author and Title

Luo F, Evans K, Norelli JL, Zhang Z, Peace C. Prospects for achieving durable disease resistance with elite fruit quality in apple breeding. Tree Genet Genomes. 2020a:16:1-14. https://doi.org/10.1007/s11295-020-1414-x

Luo F, Norelli JL, Howard NP, Wisniewski M, Flachowsky H, Hanke MV, Peace C. Introgressing blue mold resistance into elite apple germplasm by rapid cycle breeding and foreground and background DNA-informed selection. Tree Genet Genomes. 2020b:16:1-15. https://doi.org/10.1007/s11295-020-1419-5

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 2017:33(4):574-576. https://doi.org/10.1093/bioinformatics/btw663
Google Scholar: Author Only Title Only Author and Title

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018:14(1):e1005944. https://doi.org/10.1371/journal.pcbi.1005944
Google Scholar: Author Only Title Only Author and Title

McDonald BA, Stukenbrock EH. Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. Phil Trans R Soc B 2016:371(1709):20160026. https://doi.org/10.1098/rstb.2016.0026
Google Scholar: Author Only Title Only Author and Title

Morales-Cruz A, Aguirre-Liguori J, Massonnet M, Minio A, Zaccheo M, Cochetel N, Walker A, Riaz S, Zhou Y, Cantu D, et al. Multigenic resistance to Xylella fastidiosa in wild grapes (Vitis sps.) and its implications within a changing climate. Commun Biol. 2023:6(1):580. https://doi.org/10.1038/s42003-023-04938-4
Google Scholar: Author Only Title Only Author and Title

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:32(1):268-274. https://doi.org/10.1093/molbev/msu300
Google Scholar: Author Only Title Only Author and Title

Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. The complete sequence of a human genome. Science. 2022:376(6588):44-53. https://doi.org/10.1126/science.abj6987
Google Scholar: Author Only Title Only Author and Title

O'Rourke D. Economic importance of the world apple industry. The Apple Genome. 2021:1-18. https://doi.org/10.1007/978-3-030-

74682-7_1
Google Scholar: Author Only Title Only Author and Title

Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019:20(1):1-18. https://doi.org/10.1186/s13059-019-1905-y
Google Scholar: Author Only Title Only Author and Title

Peil A, Garcia-Libreros T, Richter K, Trognitz F, Trognitz B, Hanke MV, Flachowsky H. Strong evidence for a fire blight resistance gene of Malus robusta located on linkage group 3. Plant Breeding. 2007:126(5):470-475. https://doi.org/10.1111/j.1439-0523.2007.01408.x
Google Scholar: Author Only Title Only Author and Title

Perazzolli M, Malacarne G, Baldo A, Righetti L, Bailey A, Fontana P, Velasco R, Malnoy M. Characterization of resistance gene analogues (RGAs) in apple (Malus × domestica Borkh.) and their evolutionary history of the Rosaceae family. PloS one. 2014:9(2):e83844. https://doi.org/10.1371/journal.pone.0083844
Google Scholar: Author Only Title Only Author and Title

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 2015:33(3):290-295. https://doi.org/10.1038/nbt.3122
Google Scholar: Author Only Title Only Author and Title

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007:81(3):559-575. https://doi.org/10.1086/519795
Google Scholar: Author Only Title Only Author and Title

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:26(6):841-842. https://doi.org/10.1093/bioinformatics/btq033
Google Scholar: Author Only Title Only Author and Title

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11(1):1432. https://doi.org/10.1038/s41467-020-14998-3
Google Scholar: Author Only Title Only Author and Title

Schnable PS, Springer NM. Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol. 2013:64:71-88. https://doi.org/10.1146/annurev-arplant-042110-103827
Google Scholar: Author Only Title Only Author and Title

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018:15(6):461-468. https://doi.org/10.1038/s41592-018-0001-7
Google Scholar: Author Only Title Only Author and Title

Shang L, He W, Wang T, Yang Y, Xu Q, Zhao X, Yang L, Zhang H, Li X, Lv Y, et al. A complete assembly of the rice Nipponbare reference genome. Mol Plant. 2023:16(8):1232-1236. https://doi.org/10.1016/j.molp.2023.08.003
Google Scholar: Author Only Title Only Author and Title

Shang L, Li X, He H, Yuan Q, Song Y, Wei Z, Lin H, Hu M, Zhao F, Zhang C, et al. A super pan-genomic landscape of rice. Cell Res. 2022:32(10):878-896. https://doi.org/10.1038/s41422-022-00685-z
Google Scholar: Author Only Title Only Author and Title

Sherman RM, Salzberg SL. Pan-genomics in the human genome era. Nat Rev Genet. 2020:21(4):243-254. https://doi.org/10.1038/s41576-020-0210-7
Google Scholar: Author Only Title Only Author and Title

Shi X, Cao S, Wang X, Huang S, Wang Y, Liu Z, Liu W, Leng X, Peng Y, Wang N, et al. The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. Hortic Res. 2023:10(05):uhad061. https://doi.org/10.1093/hr/uhad061
Google Scholar: Author Only Title Only Author and Title

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. Bioinformatics. 2021:37(12):1639-1643. https://doi.org/10.1093/bioinformatics/btaa1016
Google Scholar: Author Only Title Only Author and Title

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31(19):3210-3212. https://doi.org/10.1093/bioinformatics/btv351
Google Scholar: Author Only Title Only Author and Title

Singh J, Sun M, Cannon SB, Wu J, Khan A. An accumulation of genetic variation and selection across the disease-related genes during apple domestication. Tree Genet Genomes. 2021:17:1-11. https://doi.org/10.1007/s11295-021-01510-1

Google Scholar: Author Only Title Only Author and Title

Song JM, Xie WZ, Wang S, Guo YX, Koo DH, Kudrna D, Gong C, Huang Y, Feng JW, Zhang W, et al. Two gap-free reference genomes and a global view of the centromere architecture in rice. Mol Plant. 2021:14(10):1757-1767. https://doi.org/10.1016/j.molp.2021.06.018
Google Scholar: Author Only Title Only Author and Title

Steuernagel B, Witek K, Krattinger SG, Ramirez-Gonzalez RH, Schoonbeek HJ, Yu G, Baggs E, Witek AI, Yadav I, Krasileva KV, et al. The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. Plant Physiol. 2020:183(2):468-482. https://doi.org/10.1104/pp.19.01273
Google Scholar: Author Only Title Only Author and Title

Sun M, Yao C, Shu Q, He Y, Chen G, Yang G, Xu S, Liu Y, Xue Z, Wu J. Telomere-to-telomere pear (Pyrus pyrifolia) reference genome reveals segmental and whole genome duplication driving genome evolution. Hortic Res. 2023:10(11):uhad201. https://doi.org/10.1093/hr/uhad201
Google Scholar: Author Only Title Only Author and Title

Sun M, Zhang M, Singh J, Song B, Tang Z, Liu Y, Wang R, Qin M, Li J, Khan A, et al. Contrasting genetic variation and positive selection followed the divergence of NBS-encoding genes in Asian and European pears. BMC genomics. 2020a:21:1-16. https://doi.org/10.1186/s12864-020-07226-1

Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, Khan A, Ban S, Xu K, Cheng L, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. Nat Genet. 2020b:52(12):1423-1432. https://doi.org/10.1038/s41588-020-00723-9
Google Scholar: Author Only Title Only Author and Title

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021:38(7):3022-3027. https://doi.org/10.1093/molbev/msab120
Google Scholar: Author Only Title Only Author and Title

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, et al. Genome evolution and diversity of wild and cultivated potatoes. Nature. 2022:606(7914):535-541. https://doi.org/10.1038/s41586-022-04822-x
Google Scholar: Author Only Title Only Author and Title

Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J, et al. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 2015:16:1-15. https://doi.org/10.1186/s13059-014-0573-1
Google Scholar: Author Only Title Only Author and Title

Van de Weyer A, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JD, Dangl JL, Weigel D, Bemm F, et al. A species-wide inventory of NLR genes and alleles in Arabidopsis thaliana. Cell. 2019:178(5):1260-1272. https://doi.org/10.1016/j.cell.2019.07.038
Google Scholar: Author Only Title Only Author and Title

Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 5Gs for crop genetic improvement. Curr opin plant biol. 2020:56:190-196. https://doi.org/10.1016/j.pbi.2019.12.004
Google Scholar: Author Only Title Only Author and Title

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. The genome of the domesticated apple (Malus × domestica Borkh.). Nat Genet. 2010:42(10):833-839. https://doi.org/10.1038/ng.654
Google Scholar: Author Only Title Only Author and Title

Wang N, Cao S, Liu Z, Xiao H, Hu J, Xu X, Chen P, Ma Z, Ye J, Chai L, et al. Genomic conservation of crop wild relatives: A case study of citrus. Plos Genet. 2023:19: e1010811. https://doi.org/10.1371/journal.pgen.1010811
Google Scholar: Author Only Title Only Author and Title

Wang T, Duan S, Xu C, Wang Y, Zhang X, Xu X, Chen L, Han Z, Wu T, et al. Pan-genome analysis of 13 Malus accessions reveals structural and sequence variations associated with fruit traits. Nat Commun. 2023:14(1):7377. https://doi.org/10.1038/s41467-023-43270-7
Google Scholar: Author Only Title Only Author and Title

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012:40(7):e49-e49. https://doi.org/10.1093/nar/gkr1293
Google Scholar: Author Only Title Only Author and Title

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The innovation. 2021:2(3). https://doi.org/10.1016/j.xinn.2021.100141
Google Scholar: Author Only Title Only Author and Title

Xiao H, Liu Z, Wang N, Long Q, Cao S, Huang G, Liu W, Peng Y, Riaz S, Walker AM, et al. Adaptive and maladaptive introgression in grapevine domestication. Proc Nat Acad Sci. 2023:120(24):e2222041120. https://doi.org/10.1073/pnas.2222041120

Google Scholar: Author Only Title Only Author and Title

Xie D, Xu Y, Wang J, Liu W, Zhou Q, Luo S, Huang W, He X, Li Q, Peng Q, et al. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun. 2019:10(1):5158. https://doi.org/10.1038/s41467-019-13185-3
Google Scholar: Author Only Title Only Author and Title

Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen JQ, Hartl DL, Tian D. Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Nat Acad Sci. 2013:110(46):18572-18577. https://doi.org/10.1073/pnas.1318211110
Google Scholar: Author Only Title Only Author and Title

Yang T, Liu R, Luo Y, Hu S, Wang D, Wang C, Pandey MK, Ge S, Xu Q, Li N, et al. Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. Nat Genet. 2022:54(10):1553-1563. https://doi.org/10.1038/s41588-022-01172-2
Google Scholar: Author Only Title Only Author and Title

Yue J, Chen Q, Wang Y, Zhang L, Ye C, Wang X, Cao S, Lin Y, Huang W, Xian H, et al. Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit Actinidia chinensis. Hortic Res. 2023:10(2):uhac264. https://doi.org/10.1093/hr/uhac264
Google Scholar: Author Only Title Only Author and Title

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun. 2019a:10(1):1494. https://doi.org/10.1038/s41467-019-09518-x
Google Scholar: Author Only Title Only Author and Title

Zhang Q, Wang Y, Wei H, Fan W, Xu C, Li T. CCR-NB-LRR proteins MdRNL2 and MdRNL6 interact physically to confer broad-spectrum fungal resistance in apple (Malus × domestica). Plant J. 2021:108(5):1522-1538. https://doi.org/10.1111/tpj.15526
Google Scholar: Author Only Title Only Author and Title

Zhang Y, Guo M, Shen J, Song X, Dong S, Wen Y, Yuan X, Guo P. Comparative genomics analysis in grass species reveals two distinct evolutionary strategies adopted by R genes. Sci Rep. 2019b:9(1):10735. https://doi.org/10.1038/s41598-019-47121-8
Google Scholar: Author Only Title Only Author and Title

Zheng W, Shen F, Wang W, Wu B, Wang X, Xiao C, Tian Z, Yang X, Yang J, Wang Y, et al. Quantitative trait loci-based genomics-assisted prediction for the degree of apple fruit cover color. The Plant Genome. 2020:13(3):e20047. https://doi.org/10.1002/tpg2.20047
Google Scholar: Author Only Title Only Author and Title

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T, Cantu D, Gaut BS. The population genetics of structural variants in grapevine domestication. Nature plants. 2019:5(9):965-979. https://doi.org/10.1038/s41477-019-0507-8
Google Scholar: Author Only Title Only Author and Title

Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y, Wu K, et al. Graph pangenome captures missing heritability and empowers tomato breeding. Nature. 2022:606(7914):527-534. https://doi.org/10.1038/s41586-022-04808-9
Google Scholar: Author Only Title Only Author and Title