

RESEARCH

Open Access



Genomes of diverse *Actinidia* species provide insights into *cis*-regulatory motifs and genes associated with critical traits

Xiaolong Li^{1,2†}, Liuqing Huo^{1,2†}, Xinyi Li^{1,2†}, Chaofan Zhang³, Miaofeng Gu^{1,2}, Jialu Fan⁴, Changbin Xu^{1,2}, Jinli Gong^{1,2}, Xiaoli Hu^{1,2}, Yi Zheng⁴ and Xuepeng Sun^{1,2*}

Abstract

Background Kiwifruit, belonging to the genus *Actinidia*, represents a unique fruit crop characterized by its modern cultivars being genetically diverse and exhibiting remarkable variations in morphological traits and adaptability to harsh environments. However, the genetic mechanisms underlying such morphological diversity remain largely elusive.

Results We report the high-quality genomes of five *Actinidia* species, including *Actinidia longicarpa*, *A. macroperma*, *A. polygama*, *A. reticulata*, and *A. rufa*. Through comparative genomics analyses, we identified three whole genome duplication events shared by the *Actinidia* genus and uncovered rapidly evolving gene families implicated in the development of characteristic kiwifruit traits, including vitamin C (VC) content and fruit hairiness. A range of structural variations were identified, potentially contributing to the phenotypic diversity in kiwifruit. Notably, phylogenomic analyses revealed 76 *cis*-regulatory elements within the *Actinidia* genus, predominantly associated with stress responses, metabolic processes, and development. Among these, five motifs did not exhibit similarity to known plant motifs, suggesting the presence of possible novel *cis*-regulatory elements in kiwifruit. Construction of a pan-genome encompassing the nine *Actinidia* species facilitated the identification of gene *DTZ79_23g14810* specific to species exhibiting extraordinarily high VC content. Expression of *DTZ79_23g14810* is significantly correlated with the dynamics of VC concentration, and its overexpression in the transgenic roots of kiwifruit plants resulted in increased VC content.

Conclusions Collectively, the genomes and pan-genome of diverse *Actinidia* species not only enhance our understanding of fruit development but also provide a valuable genomic resource for facilitating the genome-based breeding of kiwifruit.

Keywords Kiwifruits, Comparative genomics, Pan-genome, Conserved elements, Vitamin C

[†]Xiaolong Li, Liuqing Huo and Xinyi Li contributed equally to this work.

*Correspondence:

Xuepeng Sun
xs57@zafu.edu.cn

Full list of author information is available at the end of the article



Background

The kiwifruit, also known as the Chinese gooseberry, originates from various *Actinidia* species native to south-western China and was introduced to New Zealand in the early twentieth century for commercial cultivation [1–3]. The *Actinidia* genus is diverse, with over 54 species and 100 taxa, exhibiting a range of fruit morphologies, including variations in skin hairiness and color, and flesh color, juiciness, texture, and taste [4]. Particularly, *Actinidia latifolia* and *A. eriantha* have been noted for their high VC content [5], while *A. macrosperma* and *A. polygama* are known for smooth-skinned fruits, in contrast to the hairy-skinned varieties [6]. This phenotypic diversity reflects the genetic variation within the *Actinidia* genomes, suggesting the necessity to develop a pan-genome that captures the entire genetic diversity of the genus.

Recent advances in sequencing and assembly technologies have led to significant progress in decoding kiwifruit genomes. The “Hongyang” cultivar of *A. chinensis* was the first to have its heterozygous genome sequenced [2]. This genome has undergone several updates [7, 8], incorporating newer sequencing technologies over time, leading to a telomere-to-telomere, gap-free genome assembly with two haplotypes fully resolved [9]. Similar advances have been achieved for other varieties and species, including complete haplotype assemblies for *A. chinensis* var. “Donghong” and *A. latifolia* var. “Kuoye” [10], and insights into sex chromosome turnovers in *A. rufa*, *A. polygama*, and *A. arguta* [11]. Recently, a graph-based pan-genome of kiwifruit, constructed from seven representative *A. chinensis* accessions, uncovered numerous genetic variations that influence fruit traits [12]. These genomic advancements enable a more in-depth exploration of genetic variation through phylogenomic analyses.

The continuous decoding of kiwifruit genomic resources is shedding light on genes linked to key agronomic traits. Research reveals a strong positive correlation between the expression of the *GGT* gene (encoding GDP-l-galactose transferase) and the *GME* gene (encoding GDP-mannose epimerase) with L-ascorbic acid (VC) accumulation in various kiwifruit germplasms, notably in *A. eriantha*. In *Arabidopsis*, overexpressing the *GGT* gene resulted in a quadrupling of L-ascorbic acid levels compared to wild types. Similarly, transient expression of *GME* and *GGT* in tobacco led to a seven-fold increase in L-ascorbic acid. These findings confirm that the *GGT* gene catalyzes L-ascorbic acid biosynthesis via the L-galactose pathway in plants, thereby enhancing L-ascorbic acid accumulation [13–15]. Further research has led to the discovery of two new GGP transcriptional activators in *A. eriantha*. Notably, *AceGGP3*, the *GGP* with the highest expression in *A. eriantha* fruit, gets

activated by the transcription factor *AceMYBS1*, which binds to its promoter. Experiments involving overexpression and gene editing have demonstrated that *AceMYBS1* significantly elevates L-ascorbic acid levels. Another transcription factor, *AceGBF3*, has also been found to increase L-ascorbic acid content. The combined expression of *AceMYBS1* and *AceGBF3* further augments the expression of the *AceGGP3* gene [5]. Additionally, in the species *A. chinensis* “Donghong” and *A. latifolia* “Kuoye”, the transcription factor ERF098 has been identified as a potential regulator of L-ascorbic acid metabolism through comparative genomics and transcriptome sequencing studies [10]. Similarly, the genes *AcAO1* and *AcAPX2*, potentially involved in L-ascorbic acid synthesis, were identified. Overexpressing these two genes in tobacco resulted in a decrease in total L-ascorbic acid content. Furthermore, the MYB transcription factor *AcMYB123-2* has been found to inhibit L-ascorbic acid accumulation by activating the promoter of the *AcAO1* gene [16].

Despite significant efforts to decode the genes involved in L-ascorbic acid biosynthesis in kiwifruits, the molecular mechanisms underlying the diverse and unique traits across different *Actinidia* germplasm remain largely unexplored. The complexity of the genome, characterized by high heterozygosity and polyploidy, has left most of the 54 *Actinidia* species genetically uncharacterized. To further enrich the genomic resources for kiwifruit, this study assembled high-quality genomes of five *Actinidia* species. Comparative genomics analyses revealed conserved motifs and rapidly evolving gene families within *Actinidia* species. Additionally, the construction of a pan-genome facilitated the discovery of species-specific gene families in *A. latifolia* and *A. eriantha* that are involved in L-ascorbic acid biosynthesis, with gene expression levels significantly correlating with vitamin C content during fruit development. These findings enhance our understanding of the genetic basis for key agronomic traits in kiwifruit and provide a solid foundation for genome-based breeding programs for fruit improvement.

Results

De novo genome assemblies and annotation of five *Actinidia* species

Genomic sequencing and assembly were completed for the five *Actinidia* species, including *A. longicarpa*, *A. macrosperma*, *A. polygama*, *A. reticulata*, and *A. rufa*. These included four diploid species and one autotetraploid species (Additional file 1: Fig. S1; Additional file 2: Table S1). The heterozygosity rate of *Actinidia* species estimated based on the *k*-mer analysis of Illumina short reads showed significant variability, reaching as high as 10-fold difference across species (Additional file 1: Fig.

S1). The resulting monoploid genome assemblies varied in size from 579.18 Mb to 632.47 Mb, with the contig N50 lengths extending from 11.18 to 19.16 Mb, demonstrating their high continuity (Additional file 2: Table S1). On average, 99.62% of these assemblies were anchored and oriented onto 29 pseudo-chromosomes, following a reference-based scaffolding with the *A. chinensis* genome serving as the reference [9]. We predicted 43,697 to 47,228 protein-coding genes in the monoploid genome of the five kiwifruit species (Additional file 2: Table S1). BUSCO [17] analysis revealed that these genes encompass, on average, 96.32% of the 1614 broadly conserved plant genes, with *A. macrosperma* achieving 97.40%, a number comparable to the completeness of the two gap-free kiwifruit genomes reported recently [10], again indicating a high level of genome completeness for these species (Additional file 2: Table S1). These newly assembled genomes as well as their annotations have been deposited in our customized kiwifruit genome database (<http://actinidiabase.moilab.net>).

Analysis of genome collinearity revealed extensive syntenic relationship among the five kiwifruit genomes (Fig. 1a). The percentage of repetitive sequences in the genomes ranged from 36.83% in *A. macrosperma* to 43.13% in *A. polygama* (Additional file 2: Table S1). The most common repetitive sequences were long terminal repeat retrotransposons (LTR-RTs), with an average of 26.71% across the genomes. Of these, *Copia* and *Gypsy* elements comprised 10.65% and 9.66%, respectively (Additional file 2: Table S2). We detected a recent surge in the abundance of LTR-RTs in all five species (Additional file 1: Fig. S2). Notably, two species (*A. latifolia* and *A. eriantha*) with exceptionally high VC content (>200 mg/100 g FW) [5] showed significant differences in the number of intact LTR-RT elements. The *A. latifolia* had a strikingly high count of 4180 intact LTR-RTs, in contrast to *A. eriantha* with only 488 intact elements. This variation in LTR-RT abundance across kiwifruit genomes hints at a possible connection between genome flexibility and trait evolution, and offers promising loci for investigating the molecular mechanisms behind unique characteristics in various kiwifruit species.

Phylogeny and whole genome duplications

To investigate the evolutionary relationships within the Actinidia genus, we conducted a comparative analysis of their protein-coding genes against selected representative species, including tea plant (*Camellia sinensis*), *Arabidopsis thaliana*, and rice (*Oryza sativa*). This analysis identified 35,433 ortholog groups, among which 570 conserved single-copy gene families were used to construct a phylogenetic tree for the 12 species. The tree corroborated prior findings [6] and classified Actinidia species into two groups based on fruit skin hairiness: smooth-skinned fruit (SSF) including *A. macrosperma* and *A. polygama* and hairy-skinned fruit (HSF) comprising the other species (Fig. 1b). We employed three calibration points from TimeTree [18] to estimate speciation times among these species: the divergence of Arabidopsis and rice (160 MYA, confidence interval [CI]=142.1–163.5 MYA), Actinidia genus and *Camellia sinensis* (94 MYA, CI=82.8–106.0 MYA), and *A. polygama* and *A. macrosperma* (11.7 MYA). Our findings suggest that the SSF and HSF groups diverged around 12.91 MYA, while the HSF groups further split into two clades (C1 and C2) around 8.41 MYA. The ancestor of *A. latifolia* and *A. eriantha* diverged from its common ancestor with *A. reticulata* around 7.85 MYA, aligning with previous findings [10]. Moreover, the *A. chinensis* clustered closest with the diploid variety of *A. deliciosa* [19], which diverged around 6.76 MYA (Fig. 1b). Additionally, we inferred the demographic history of five Actinidia species using the multiple sequentially Markovian coalescent (MSMC) algorithm [20] based on the whole genome sequencing data. The results showed that a common and protracted bottleneck occurred approximately 0.5–20 million years ago, after which the population sizes rebound for all of these species (Fig. 1c).

Kiwifruits are known to have experienced whole genome duplications (WGDs) [2, 21]. To assess the occurrence of duplication events in the newly sequenced genomes, we calculated the synonymous substitution rates (K_s values) for paralog pairs in kiwifruit genomes. The analysis of these parameters revealed three significant WGDs shared across all nine kiwifruit species

(See figure on next page.)

Fig. 1 Comparative genomic analyses of five de novo assembled Actinidia genomes. **a** Genome collinearity analysis revealed extensive syntenic relationship among the five kiwifruit genomes. The colored segments represent different genomes. Grey lines represent collinear blocks. **b** Molecular dating and gene family evolution analysis of the nine Actinidia species and the three selected outgroups. The maximum likelihood phylogeny was constructed using conserved single-copy orthologs. The speciation times among different plant lineages are labeled in the bottom with parentheses. The numbers on each branch represent the expanded (red) or contracted (blue) gene families. SSF refers to smooth-skinned fruit group, including *A. macrosperma* and *A. polygama*; HSF refers to hairy-skinned fruit comprising the other Actinidia species. MYA refers to million years ago. "J" indicates Jurassic, "K" indicates Cretaceous, "Pg" indicates Paleogene, "N" indicates Neogene, and "Q" indicates Quaternary. **c** Population demographic history of five newly assembled kiwifruit species. **d** The distribution of synonymous substitution rate (K_s) of paralogs in each of the nine kiwifruit species. Different colors represent different species

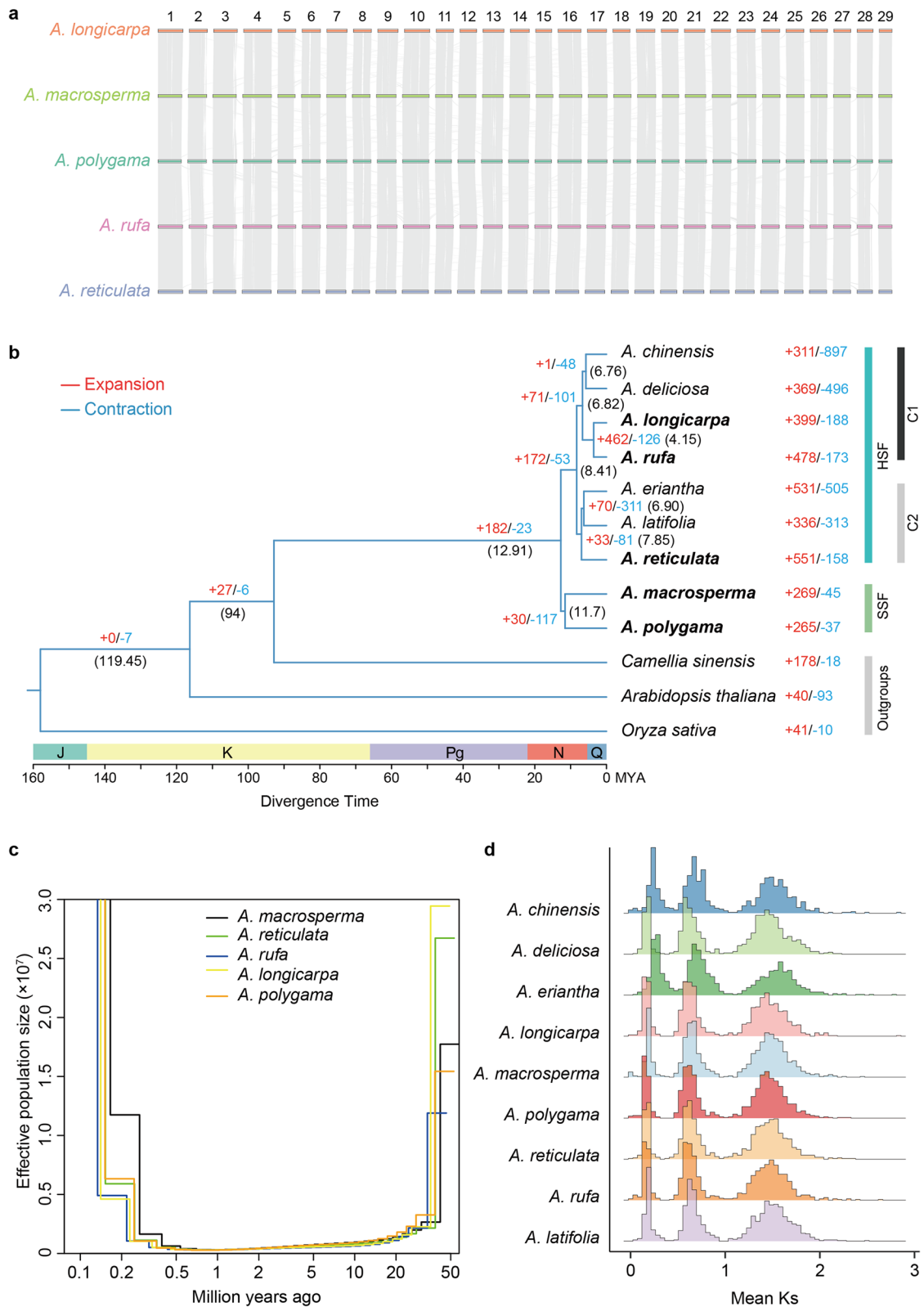


Fig. 1 (See legend on previous page.)

studied (Fig. 1d). Thus, these duplication events, both of genes and repetitive regions, may have been instrumental in driving the evolution and expansion of kiwifruit genomes.

Structure variations and conserved noncoding sequences

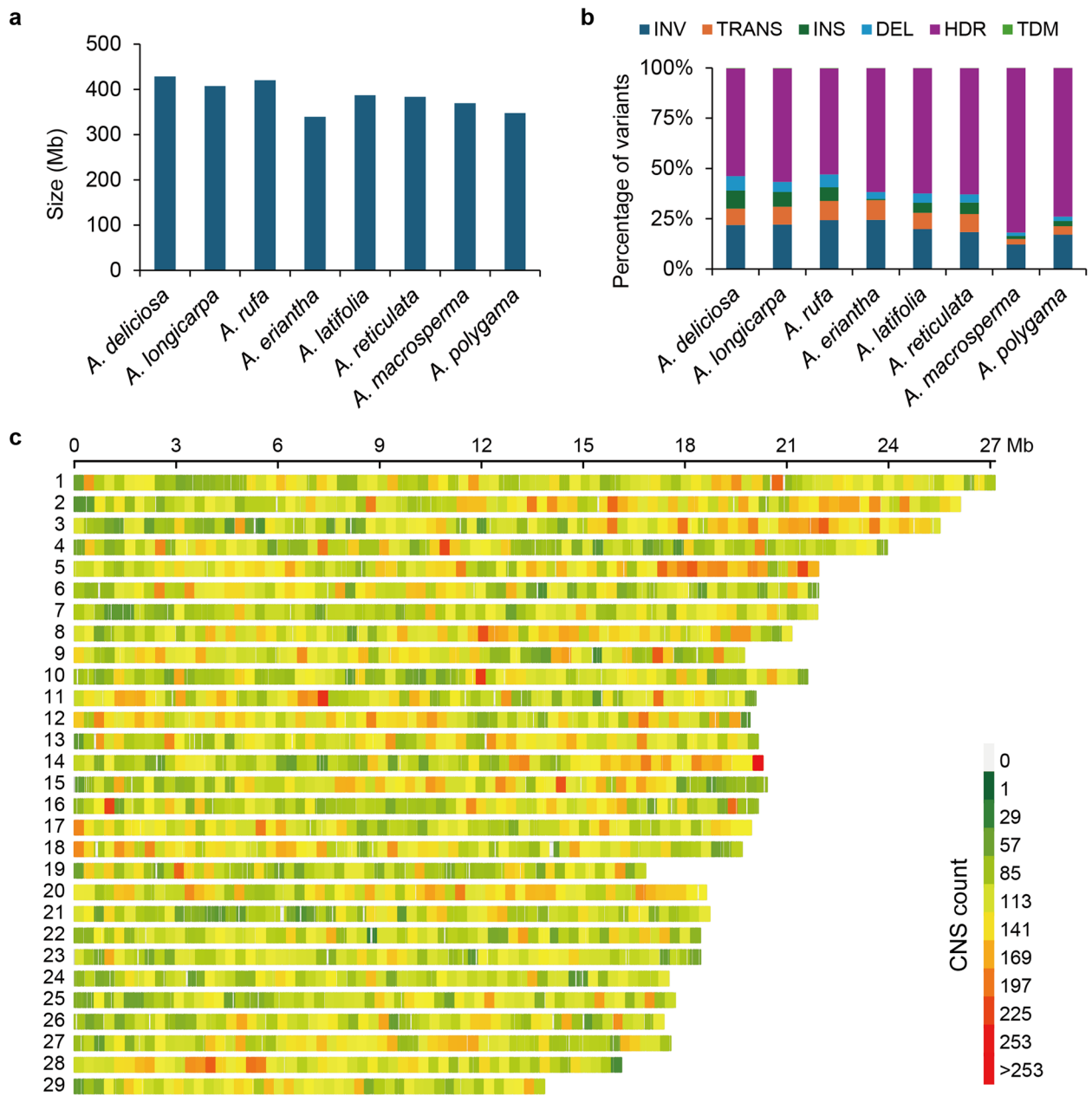
Taking the *A. chinensis* genome as a reference, we identified structure variations (SVs), including insertions (INs), deletions (DELs), inversions (INVs), translocations (TRANSs), high-divergence regions (HDRs), and tandem repeats (TDMs), from the newly assembled genomes as well as previously published three kiwifruit genomes from *A. deliciosa*, *A. latifolia*, and *A. eriantha* [10, 19, 21] using the assembly-based mapping method, which enabled us to identify more accurate structure variations (Fig. 2a, b; Additional file 2: Table S3). Our results showed that syntenic regions were more extensive in species closer to the reference species, while the HDRs predominated in variable genomic regions (Fig. 2a, b). A total of 222 species-specific SVs longer than 50 bp were identified in the VC enriched *A. latifolia* and *A. eriantha*, affecting 174 genic regions and 9 promoter regions (2 kb upstream TSS) (Additional file 2: Table S4–6). Interestingly, 28 genes associated with SVs were found to show significant correlation ($r > 0.7$ and P value < 0.01) with VC concentration as verified by integrative analysis of recently published fruit transcriptomic and metabolomic data at various developmental and ripening stages [16] (Additional file 2: Table S7). Among these, a MYB transcription factor (Actinidia07194) involved in regulation of alcohol dehydrogenase activity was identified in the HDR region. Alcohol dehydrogenase has been reported to regulate L-ascorbic acid synthesis and influence cold tolerance in plants [22–24]. Furthermore, the hair on the fruit skin is a key feature distinguishing smooth-skinned from hairy-skinned kiwifruit accessions. In this study, the *A. polygama* and *A. macrosperma* are the only two kiwifruit species with smooth-skinned fruits. We identified 463 specific SVs in the smooth-skinned fruit group, affecting 323 genic regions and 26 promoter regions (Additional file 2: Table S8–10). Among these, a WUSCHEL-related homeobox (WOX) gene associated with an inverted translocated region was identified, which has been reported to regulate leaf hair formation in rice [25]. In conclusion, these candidate SVs provide valuable genomic resources for exploring molecular mechanisms underlying the morphological variations in Actinidia species.

Additionally, the nine Actinidia genomes were analyzed to identify conserved noncoding sequences (CNSs). To provide a broader phylogenetic context, *Camellia sinensis*, a member of the Theaceae family that is sister to Actinidiaceae, was included as an outgroup. From the

genome alignments of these nine Actinidia species, a total of 227,786 conserved regions were identified. These regions showed a size range from 7 to 33,978 bp, with an average length of 971.83 bp under a P value cutoff of 0.05 (Fig. 2c). Notably, these conserved regions collectively accounted for 37.96% of the *A. chinensis* genome. Further analysis revealed that out of these conserved regions, 40,318 CNSs were found within the upstream 2-kb region of genes. This subset represented 3.28% of the *A. chinensis* genome, encompassing a total sequence length of 19.14 Mb. In a similar distribution, 39,836 CNSs were located in the downstream 2-kb region of genes, occupying 3.27% of the *A. chinensis* genome, and spanning a total sequence length of 19.11 Mb. This comprehensive identification of CNSs across multiple Actinidia species, along with the comparative analysis with *C. sinensis*, provides valuable insights into the evolutionary conservation and potential functional roles of these genomic regions.

Identification of conserved and novel cis-regulatory motifs

Since CNS analysis does not directly provide information for regulatory motifs, we employed word-based algorithms [26] to detect conserved and novel *cis*-regulatory elements, incorporating phylogenetic footprinting across nine Actinidia species. An exhaustive search of all nucleotide combinations, including different level of degeneracy for motif lengths between 6 and 8 bp, resulted in the identification of 1,312,474 potential motifs within 30,289 gene families. Among these, 8278 motifs were significantly enriched comparing with random sequences. Redundancy filtering yielded a final set of 76 motifs, with occurrences ranging from 27 to 2000 genes (Additional file 1: Fig. S3). Genes sharing the same motifs are hypothesized to be co-regulated and implicated in similar biological processes. To evaluate the functional similarity of gene sets containing identical motifs, we conducted Gene Ontology (GO) enrichment analysis. Gene sets corresponding to 64 out of 76 motifs were enriched for at least one biological function (adjusted $P < 0.05$), predominantly associated with stress response, metabolic processes, and development (Fig. 3). Motif position analysis revealed that 72% of motifs were located within 1 kb of the transcription start site (TSS) (Fig. 4a). Further validation of the motifs was pursued through chromatin accessibility assays using ATAC-Seq [27], demonstrating significantly enhanced chromatin accessibility in motif regions across all examined tissues (Fig. 4b). Given the role of histone modifications as indicators of active promoters and gene transcription, we performed CUT&Tag analysis of active histone marks H3K4me1 and H3K9ac in kiwifruit plant leaves. Both marks exhibited enrichment towards the center of motif regions (Fig. 4c), collectively supporting



CNS Distribution

Fig. 2 Identification of structure variations (SVs) and conserved noncoding sequences (CNSs). **a** Cumulative length of genome sequences that can be aligned to the *A. chinensis* genome in each of the eight kiwifruit species. **b** The proportion of different type of SVs, including insertions (INSS), deletions (DELS), inversions (INVs), translocations (TRANSs), high-divergence regions (HDRs), and tandem repeats (TDMs). **c** The genomic distribution of CNSs on 29 chromosomes using *A. chinensis* genome as a reference. Red blocks indicated high density, and green indicated low density

the authenticity of the identified motifs in Actinidia species.

Subsequently, we annotated these motifs by comparison with experimentally verified motifs in *A. thaliana*, sourced from a DAP-seq experiment [28]. Of the 76 predicted motifs, 71 exhibited significant matches to

known Arabidopsis motifs, including those recognized by transcription factors such as TCP16, EIN3, ERF10, and AREB3 (Fig. 3). Notably, the remaining 5 motifs (i.e., HWGGCCCA, CKCTCKAG, CCMAABGG, GCC CAGCC, CCCGGCCC), which did not correspond to any known motifs, may represent novel regulatory

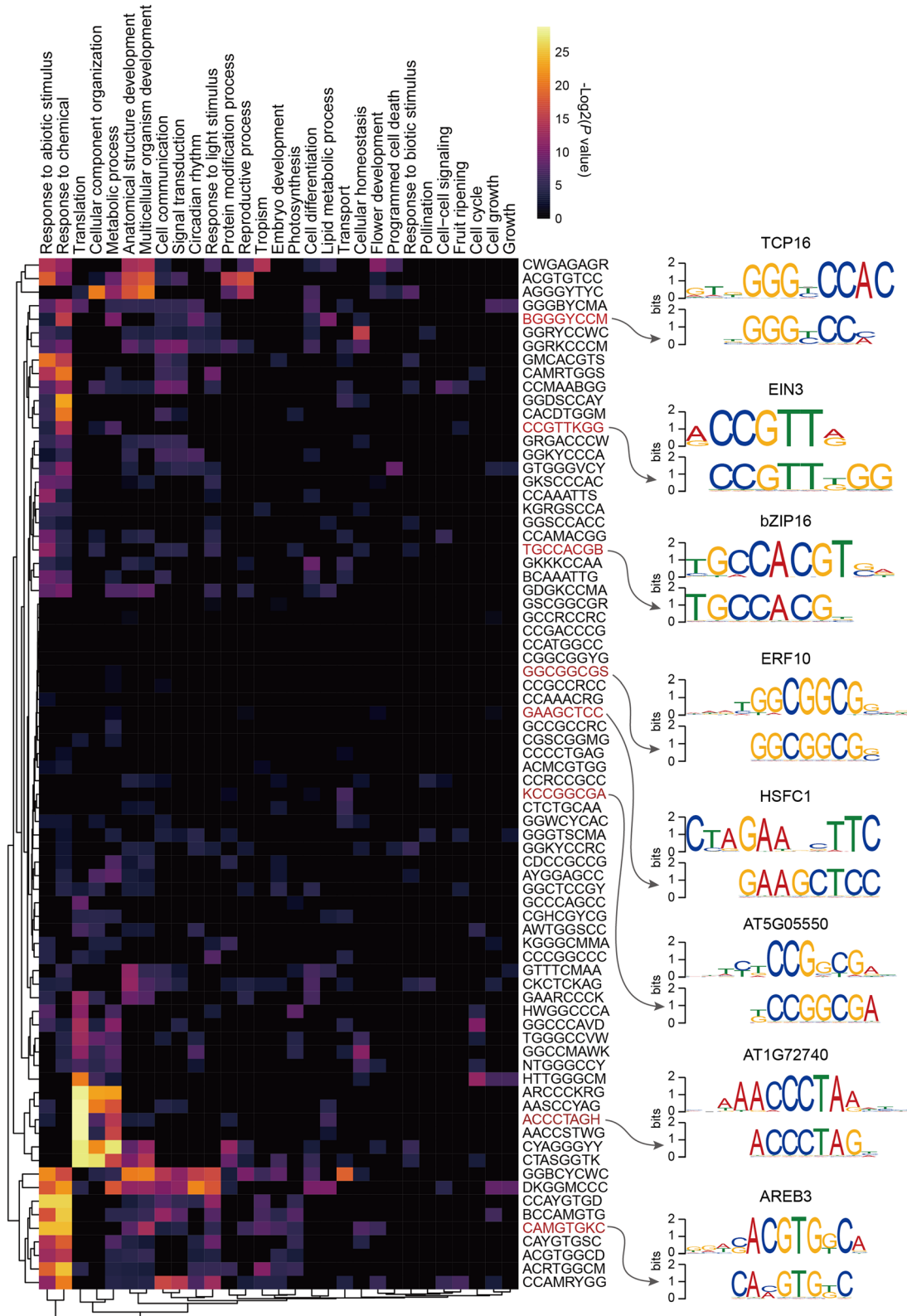


Fig. 3 Discovery and the function of *cis*-regulatory motifs. The heatmap shows the GO enrichment of potential targets of each motif. Only biological processes in the GO terms were plotted. Motifs marked with red were chosen as representatives that showing significant similarity with known motifs in Arabidopsis. Comparison of discovered motifs in Actinidia and known motifs of Arabidopsis transcription factors was plotted on the right

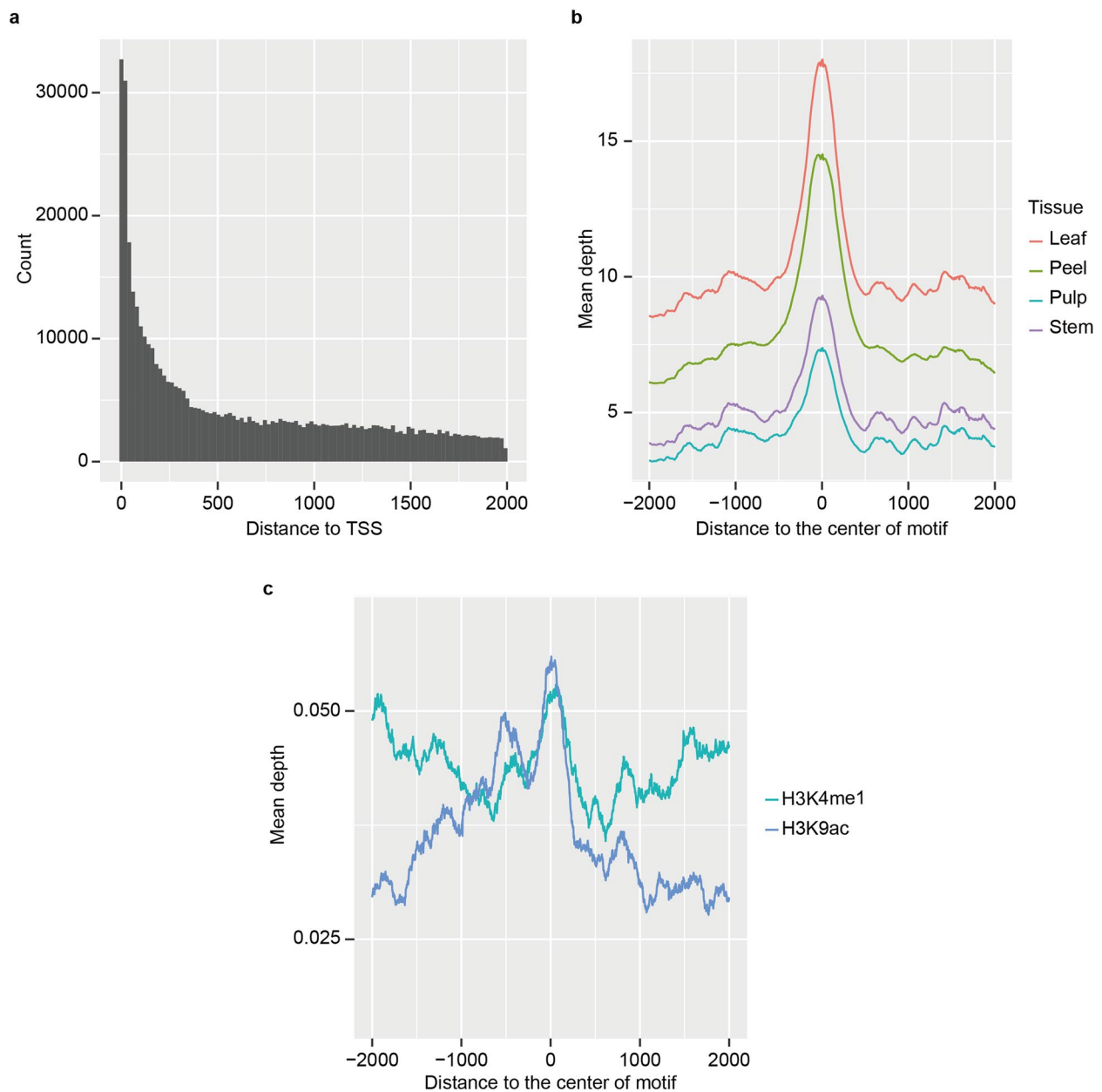


Fig. 4 Genomic features of the identified *cis*-regulatory motifs in Actinidia. **a** A histogram showing the distance distribution of identified *cis*-regulatory motifs to its downstream gene. **b** The chromatin accessibility of genomic regions encompassing the identified motifs. Chromatin accessibility was measured by ATAC-Seq. The depth of ATAC-Seq reads (y axis) is positively correlated with the accessibility of chromatin. Data from different tissue of kiwifruit plant were labeled with different colors. **c** The level of active histone modifications, H3K4me1 and H3K9ac, in the genomic regions encompassing the identified motifs. Histone modifications were measured by CUT&Tag and the depth of sequencing reads (y axis) is positively correlated with the level of corresponding epigenetic marks

elements in Actinidia species. Although their specific functions remain to be elucidated, these novel motifs constitute a valuable resource for future research into their regulatory roles in various biological processes.

Gene family evolution

We examined the expansion and contraction of gene families using the established phylogenetic tree of Actinidia species. Notably, as the kiwifruit species *A. latifolia* and *A. eriantha*, known for their exceptionally high

VC content, diverged from other kiwifruit species, we observed significant expansion ($P < 0.05$) in 70 gene families and contraction in 311 families (Fig. 1b). Functional analysis of these expanded genes revealed a predominant presence in flavonoid metabolic process and calcium ion homeostasis (Additional file 2: Table S11). Similarly, in the comparison of hairy-skinned and smooth-skinned (*A. polygama* and *A. macrosperma*) kiwifruit species, we found significant expansion in 30 gene families and contraction in 117 families (Fig. 1b). GO enrichment analysis indicated that the substantially contracted gene families in hairy-skinned kiwifruits were mainly associated with cell wall biogenesis (Additional file 2: Table S12).

Given that fruit hairiness is a critical trait distinguishing different groups of Actinidia species, we sought to determine whether the contraction of specific gene families might underlie these morphological variations. To address this, we identified trichome-related genes in *A. chinensis* through transcriptome profiling of the trichomes and the naked stem (with trichomes mechanically removed) (Fig. 5a). This analysis revealed 4435 genes that were upregulated and 4944 genes that were downregulated in the trichomes compared to the stem (fold change ≥ 2 and adjust $P \leq 0.01$). Consistent with the GO enrichment analysis of contracted gene families in SSF kiwifruit, the biological processes associated with upregulated genes in trichomes were predominantly associated with cell wall functions (Fig. 5b). Among these differentially expressed genes, 66 genes belonging to families that were contracted in SSF kiwifruits showed differential expression (Fig. 5c). Notably, this included the bHLH transcription factor Actinidia07314, which is absent in *A. polygama* and *A. macrosperma* but present in HSF kiwifruits (Additional file 1: Fig. S4). This observation aligns with the established role of bHLH transcription factors as key regulators of trichome biogenesis in plants [29] and implies a possible role of this gene in the development of trichomes in hairy-skinned kiwifruits.

The pan-genome of Actinidia species

We constructed a pan-genome for the Actinidia species by analyzing the nine kiwifruit genomes. Our investigation identified a total of 37,915 gene families across these nine species, with 14,675 gene families classified as core (present in all nine genomes) and 21,749 (57.36%) as dispensable (found in 2–8 species). We observed that while the total gene family count increased with the inclusion of each additional genome, the rate of increase for core and pan-gene families diminished after the inclusion of more than seven genomes (Fig. 6a). This may suggest that the nine species analyzed effectively capture the genetic diversity represented in the Actinidia pan-genome. Notably, 1491 gene families were identified as unique to one

of the nine Actinidia species, with the proportion of these unique gene families varying from 0.09% to 1.50% across different genomes. *A. eriantha* exhibited the highest number of unique gene families ($n = 570$), whereas *A. longicarpa* had the fewest ($n = 33$) (Fig. 6b, c). This comprehensive pan-genome provides a robust foundation for subsequent studies on genome evolution, genetic diversity, and the functional genomics of kiwifruits.

Focusing on the two high VC kiwifruit species, *A. latifolia* and *A. eriantha*, we identified 234 specific orthogroups. The majority of these gene families (78.63%) comprised single-copy genes from both *A. latifolia* and *A. eriantha*, with *A. eriantha* containing a higher copy number in the remaining specific gene families (Fig. 6d). Notably, one gene family of interest included the gene *Alf23g04080KY* from *A. latifolia* and two homologous copies (*DTZ79_05g11960* and *DTZ79_23g14810*) from *A. eriantha*, which are analogous to an Arabidopsis gene involved in L-ascorbic acid synthesis [30, 31]. Revisiting published RNA-seq data [32] revealed that the gene *DTZ79_23g14810* exhibited high expression levels 20 days post-anthesis (DPA), which correlated strongly ($r = 0.97$) with L-ascorbic acid concentrations during different stages of fruit development (Fig. 6e). Conversely, the gene *DTZ79_05g11960* showed minimal or no expression at these developmental stages in *A. eriantha*. To further confirm the role of *DTZ79_23g14810* in L-ascorbic acid biosynthesis, we introduced this gene into kiwifruit plant using an efficient root transformation system that we have established recently in several woody plants [33]. Our results indicated that plant transgenic roots overexpressing *DTZ79_23g14810* exhibited remarkable higher VC content compared to wild-type (WT) roots (Fig. 6f). These findings suggest that the evolution of critical traits in Actinidia species, such as VC biosynthesis, can be partially attributed to the gain or loss of specific genes or gene families.

Discussion

Kiwifruit holds a significant place in the global agricultural economy, valued for its rich VC content and forming a vital component of people's daily diets by providing fresh, delicious, and nutritious fruit to consumers worldwide. The current market demands fruits that are both esthetically pleasing and flavorful, reflecting the high standards and varied preferences of consumers. However, challenges such as climate change and the rise in plant diseases and pest infestations threaten crop production, including that of kiwifruit species. These challenges necessitate the development of superior kiwifruit cultivars that not only meet consumer preferences but are also resilient to biotic and abiotic stress, demanding more efficient and effective breeding practices. Genomic

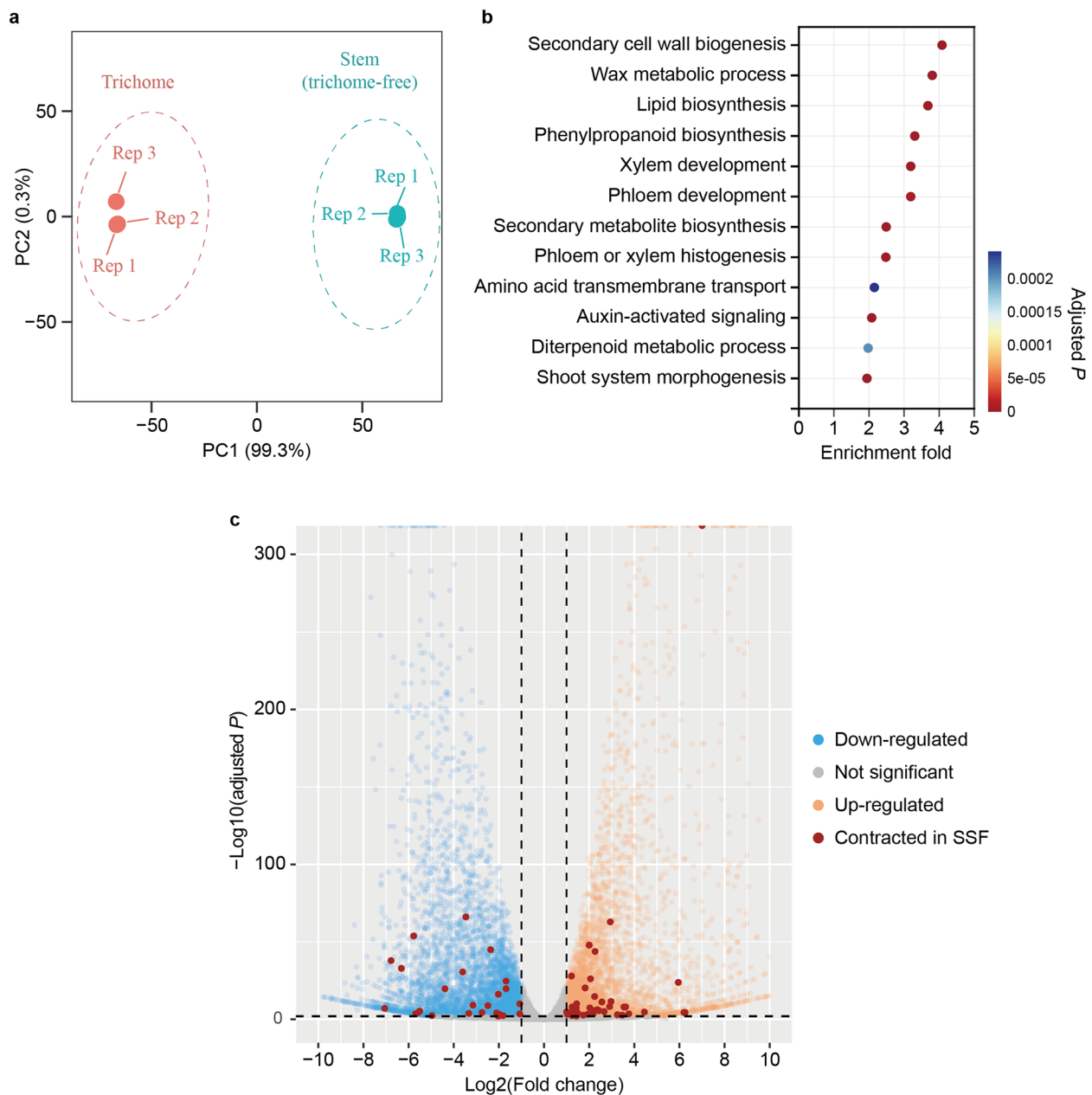


Fig. 5 RNA-Seq analysis of trichomes and naked stem (with trichomes mechanically removed) of *A. chinensis*. **a** The PCA plot of RNA-Seq data from trichomes and naked stem tissues. Red points represent the three biological replicates of trichomes, and green points represent the biological replicates of stem. **b** GO enrichment analysis of upregulated genes in trichomes. **c** The volcano plot of differentially expressed genes. Blue points represent the downregulated genes; orange points indicate upregulated genes; grey points mark the genes with no significant difference; red points represent the genes identified from contracted gene families in SSF group

research offers detailed guidance for rapid and successful plant breeding. Yet, the reliance on a single reference genome from one accession has traditionally limited our understanding of the full array of genetic variation and structural variations (SVs) present across different accessions within a species or genus. With the decreasing costs and advancements in sequencing technology, we are

entering an era of plant pan-genomics, where multiple genomes are analyzed to capture a more comprehensive genetic picture.

The introduction of five high-quality kiwifruit genomes and an integrated pan-genome from nine kiwifruit accessions marks a significant step forward for in-depth functional genomics studies within the *Actinidia* genus.

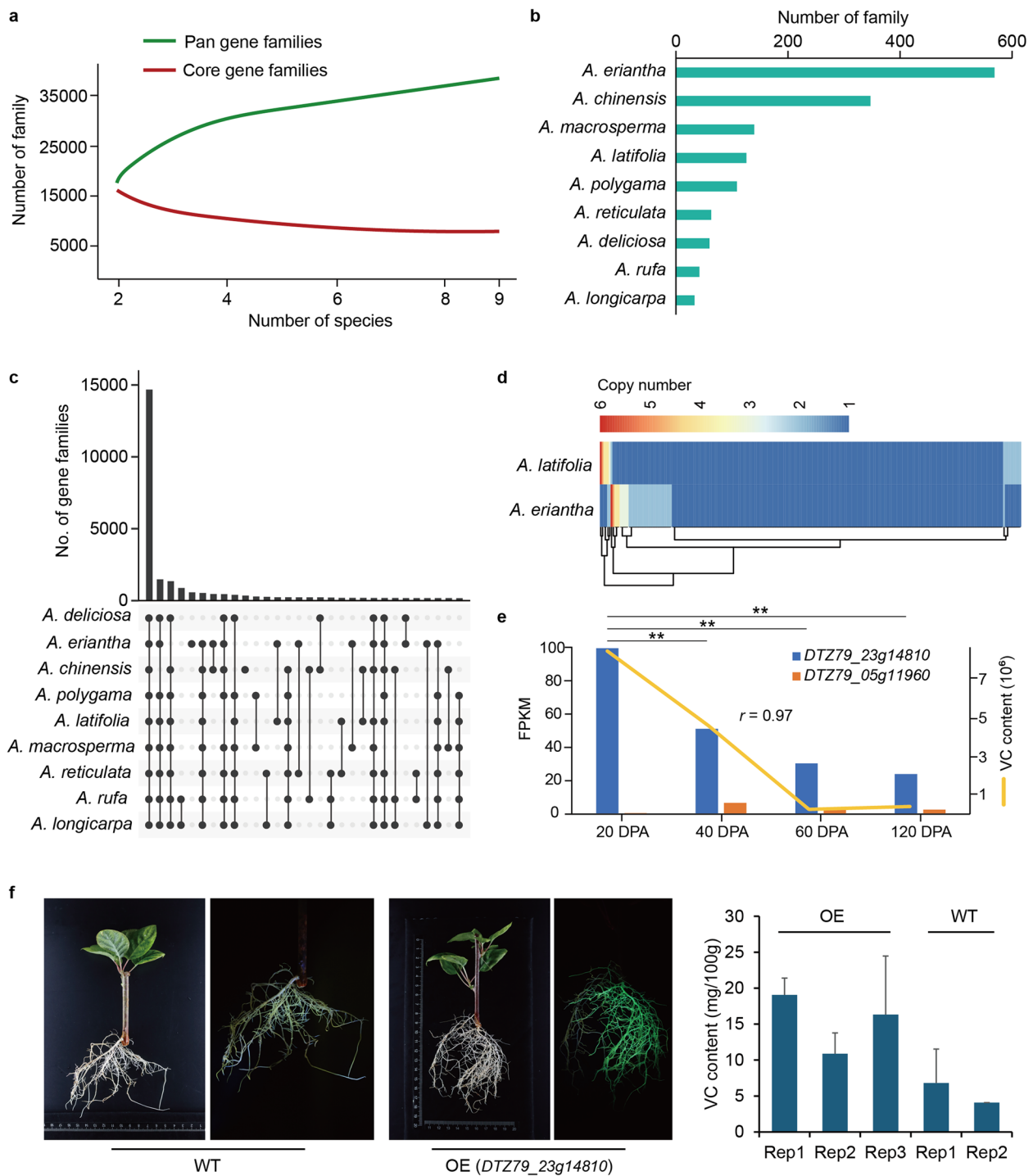


Fig. 6 Pan-genome construction and the identification of genes associated with vitamin C (VC) content. **a** Statistics of the pan and core gene clusters among the nine kiwifruit genomes. **b, c** The number of species-specific (**b**) and shared gene families (**c**). **d** The specific orthogroup identified from the two high VC content kiwifruit species, *A. latifolia* and *A. eriantha*. **e** Gene expression levels of the two paralogs (*DTZ79_05g11960* and *DTZ79_23g14810*) from *A. eriantha* and its correlation with VC concentrations during different stages of fruit development (20 DPA, 40 DPA, 60 DPA, 120 DPA). Yellow line represents the VC content. Asterisks indicate statistical significance under $P < 0.05$. **f** An efficient root transformation system was introduced to generate the kiwifruit transgenic roots overexpressing *DTZ79_23g14810*. Right panel shows the VC content of overexpressing (OE) and wild-type (WT) roots. Error bar indicates standard error of the measurements

Through comparison and phylogenomic analysis, we can now identify more complex and larger SVs and conserved and novel regulatory motifs as well as expansions and contractions in gene families linked to key evolutionary traits. This study has honed in on two distinct phenotypes: the texture (hairy or smooth) and the VC content (high or low) in fresh kiwifruit. We observed that reductions in gene families involved in critical biological processes may influence the distinct flavors of *A. latifolia* and *A. eriantha*, both of which have ultra-high VC content. Furthermore, in species such as *A. polygama* and *A. macrosperma*, which are characterized by hairy fruit, expansions in gene families related to cellular differentiation and cell wall biogenesis may contribute to the development of their hairy fruit trait. Additionally, analysis of the population demographic history revealed a prolonged bottleneck shared by various Actinidia species, a pattern also observed in other fruit crops [34]. Global climate changes, including recurrent glaciation, are well known to reduce the effective population sizes of plants [34]. Due to the dioecious nature of kiwifruit plants, their population size is particularly vulnerable to adverse environmental conditions.

The current narrow base of breeding resources has led to a significant decline in genetic diversity among modern kiwifruit cultivars, severely limiting the potential for further improvement. The genus Actinidia, comprising approximately 54 species with extensive genetic and morphological diversity [35], represents a valuable reservoir of untapped genetic potential. These species could serve as valuable breeding parents, introducing new and desirable alleles crucial for the future enhancement of kiwifruit breeding and the preservation of genetic diversity.

Conclusions

In conclusion, the release of these high-quality genomes and the construction of a pan-genome at the genus level are not only instrumental in advancing our understanding of genome evolution, genetic diversity, and functional genomics but also provide a robust foundation for precision breeding. These resources will facilitate the continued improvement of kiwifruit by enabling the introduction of novel traits and enhancing genetic resilience.

Methods

Genome sequencing and assembly of five Actinidia species

Young leaves from the mature plants of five Actinidia species, including *A. longicarpa*, *A. macrosperma*, *A. polygama*, *A. reticulata*, and *A. rufa* were collected from a germplasm garden in Xi'an, Shaanxi Province, and used for high-molecular weight (HMW) DNA extraction following the CTAB (2%) method. The ONT libraries were

prepared using the Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, Cambridge UK), following the manufacturer's guidelines. The HMW DNA was also used for preparing Illumina paired-end (PE) libraries with a mean insertion size of approximately 350 bp. The Illumina libraries were sequenced on the NovaSeq6000 system (150-bp × 2).

The ONT reads were assembled with both the NextDenovo [36] and NECAT [37] software under default parameters. The assemblies were processed to remove haplotigs with `purge_haplotigs` program [38] and polished with Illumina short reads using NextPolish [39]. The final assembly was obtained by comparing and merging of the polished NextDenovo assembly and NECAT assembly with `quickmerge` [40]. Reference-based anchoring of contigs into pseudo-chromosomes were performed using RagTag tools [41].

Repetitive elements and gene annotation

Repetitive sequences were identified using the EDTA tool [42]. These repetitive elements were used to produce a soft-masked genome for gene structure annotations with BRAKER2 [43]. To make an accurate gene model prediction, we used RNA-Seq and cross-species proteins to predict protein-coding genes separately and then incorporated both types of evidence simultaneously using the TSEBRA pipeline [44]. The predicted proteins were functionally annotated by querying against databases such as the TrEMBL (<https://www.uniprot.org/>), KEGG (<https://www.genome.jp/kegg/>), and SwissProt using BLASTP with the *E*-value threshold set to be 1E-5. The best hits from these searches were utilized to assign functions to genes. Additionally, GO categories (<http://geneontology.org/>) and InterPro entries (<https://www.ebi.ac.uk/interpro/>) were determined using InterProScan [45].

Comparative genomics analyses

Non-redundant protein sequences from the 12 species were prepared for ortholog analyses, including *A. chinensis*, *A. deliciosa*, *A. latifolia*, *A. eriantha*, *A. longicarpa*, *A. macrosperma*, *A. polygama*, *A. reticulata*, and *A. rufa*, *Camellia sinensis* (<https://ngdc.cnpc.ac.cn/search/?dbId=gwh&q=GWHASIV000000000+>), *Arabidopsis thaliana* (<http://www.arabidopsis.org/>), and *Oryza sativa* (<https://data.jgi.doe.gov/refine-download/phytozome?q=Oryza+sativa>). The orthologous gene groups were clustered by OrthoFinder v2.5.4 [46] with default parameters. Conserved single-copy genes were used to construct the species phylogeny. We used CAFE5 [47] to identify the expansion and contraction of gene families under the *P* value cutoff of 0.01. The MCMCTree program from PAML v4.9 [48] was used to estimate speciation time with the following parameters: `burnin = 200,000`, `sampfreq = 100`,

and $n_{\text{sample}} = 50,000$. The divergence times of Arabidopsis and rice (160 MYA, CI = 142.1 ~ 163.5 MYA), *A. longicarpa* and *Camellia sinensis* (94 MYA, CI = 82.8 ~ 106.0 MYA), and *A. polygama* and *A. macrosperma* (11.7 MYA) were chosen as three calibration points using the Time-Tree database (<http://timetree.org>). Paralog pairs for each of nine kiwifruit species was identified with MCScanX [49], and the WGD events of each sample were estimated from the distribution of K_s for each paralog pair.

Variation call and pan-genome construction

Genome assemblies were aligned to the *A. chinensis* “Hongyang” genome using Mummer (v4.0) [50] with the parameters set as “-l 50 -c 100 -maxmatch”. The alignments were filtered using delta-filter with parameters “-m -i 90 -l 100” according to the methods described in previous studies [10, 51]. The alignments were then used to call structure variations by using the SyRI pipeline [52]. The orthologous gene groups were clustered by OrthoFinder with default parameters [46], and the orthogroups were used to construct the pan-genome using in-house scripts.

Demographic history of Actinidia species

To infer historic changes in effective population size (N_e) of five Actinidia species, we employed the multiple sequentially Markovian coalescent (MSMC) approach [53] based on next-generation whole genome sequencing data. The MSMC, an advancement of the pairwise sequentially Markovian coalescent (PSMC) method, yields similar results to PSMC when analyzing a single individual but offers greater accuracy. We scaled the results of our demographic modeling using a substitution rate of 3.39×10^{-9} mutations per site per year [8] and assumed a generation time of 5 years.

Identification of conserved noncoding sequences (CNSs)

To identify conserved noncoding sequences (CNSs), we utilized the previously published CNSpipeline [54]. Initially, each genome was subjected to generate a pairwise whole-genome alignment (WGA) against the *A. chinensis* genome [51] using Last v2.34 [55]. Prior to alignment, simple repetitive elements in all genome sequences were masked using Tantan [56]. The alignments were generated using the lastdb argument “-uMAM8,” and lastal arguments “-p HOXD70 -e 4000 -C 4 -P 6 -m 100”. Subsequently, the split-last algorithm [57] was employed to refine the selection of orthologous sequences. These pairwise WGAs were then organized using axtChain and chainNet [58]. All pairwise alignments were combined into multiple sequence alignments using the Roast software (<http://www.bx.psu.edu/~cathy/toast-roast.tmp/README.toast-roast.html>), which merges the

alignments using a phylogenetic tree-guided approach. In the final step, we calculated the score for each base, and CNSs were identified using a P value threshold of 0.05.

Discovery of cis-regulatory motifs

Motif discovery was conducted using the BLSSpeller pipeline [26]. Briefly, the phylogeny of the nine Actinidia species was built using single-copy orthogroups. For each orthogroup, the 2-kb upstream sequences of each gene were extracted to identify potential motifs with lengths between 6 to 8 bp and a degeneration level of 3. The conservation of motifs was defined by a branch length score (parameter: blsthresholds), which was set to be “0.05,0.1,0.25,0.5,0.75,0.85,0.95”. To obtain a reliable set of motifs, we selected motifs with a recurrence score of 0.8 in at least one of the specified BLS thresholds. The final set of motifs was obtained by filtering out redundant motifs, defined as those having more than 50% overlap in targets between any two motifs.

Samples collection and RNA sequencing

The stems of *A. chinensis* “Hongyang” were rapidly frozen in liquid nitrogen. Trichomes were then scratched from the stems in liquid nitrogen and immediately used for total RNA isolation, while the trichome-free stems were similarly prepared for RNA isolation. Each tissue sample was collected in triplicates. RNA extraction was carried out using the Plant Total RNA Isolation Kit Plus (FOREGENE Co.), followed by treatment with RNase-Free DNase I. Sequencing libraries were generated using the VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina® (NR604-01/02), following the manufacturer’s recommendations. The libraries were sequenced on an Illumina NovaSeq 6000 platform.

Quality control was performed using fastp [59] to obtain cleaned data. The reference genome of *A. chinensis* was indexed, and the cleaned reads were mapped to the reference genome using the Hisat2 software [60]. Alignments were processed using featureCounts [61] to count the number of reads mapping to each gene. RPKM values were calculated using a custom Python script based on the mapped reads. Differentially expressed genes between the tissues were identified using the R package “DESeq2” [62].

Root transformation and VC content measurement

Healthy, semi-lignified branches were selected for agroinfiltration using the following procedure. Briefly, *A. rhizogenes* strain K599 carrying the overexpressing plasmid was cultured overnight in liquid LB medium supplemented with appropriate antibiotics at 28 °C until the OD_{600} value reached 0.8–1.0. The cultures were centrifuged at 6000 g for 10 min at room temperature and then

resuspended in infiltration buffer (0.05 M MES, 2 mM Na₃PO₄, 0.5% (w/v) D-glucose, and 0.1 mM acetosyringone) to a final OD₆₀₀ of 0.8. Branches of *A. valvata* were collected and cut into the segments of 5–10 cm in length using sterilized pruning shears. The base of the stems was then immersed in the *A. rhizogenes* suspension and vacuum infiltrated for approximately 30 min under standard vacuum conditions. Subsequently, the branches were inserted into sterilized vermiculite and placed in the greenhouse for cultivation, maintaining a temperature of 26 °C with a light cycle of 16 h of illumination followed by 8 h of darkness. Approximately 2–4 weeks after agro-infiltration, the success of genetic transformation was assessed by detecting the fluorescence of hairy roots using a portable excitation lamp (Luyor-3415RG, Shanghai, China). The transgenic and wild-type roots were collected and homogenized for VC content measurement using the kit purchased from Suzhou Grace Biotechnology Co., Ltd., China. (Art. No. G0201F).

Statistical analysis

Statistical analyses were computed with the R statistics software. All graphs were drawn using “ggplot2” in R package.

Abbreviations

VC	Vitamin C
LTR-RTs	Long terminal repeat retrotransposons
Ks	Synonymous substitution rate
HSF	Hairy-skinned fruit
SSF	Smooth-skinned fruit
WGD	Whole genome duplication
WOX	WUSCHEL-related homeobox
INs	Insertions
DELs	Deletions
INVs	Inversions
TRANSs	Translocations
HDRs	High-divergence regions
TDMs	Tandem repeats
GO	Gene ontology
TSS	Transcription start site
DPA	Days post-anthesis
QE	Overexpressing
WT	Wild-type
SVs	Structural variations
HMW	High-molecular weight
CI	Confidence interval
MSMC	Multiple sequentially Markovian coalescent
PSMC	Pairwise sequentially Markovian coalescent
CNSs	Conserved noncoding sequences
WGA	Whole-genome alignment
AsA	Ascorbic acid
FW	Fresh weight

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-02002-z>.

Additional file 1: Fig. S1. Estimation of the size and heterozygosity rate of kiwifruit genomes based on the *k*-mer analysis of Illumina short reads using GenomeScope. Fig. S2. Estimation of the insertion time of LTR-RTs in

different Actinidia species. Fig. S3. The number of potential targeted genes of each *cis*-regulatory motif. Fig. S4. Maximum likelihood phylogeny of genes absent in *A. macrosperma* and *A. polygama*.

Additional file 2: Table S1. Summary statistics of the sequencing data and genome assemblies. Table S2. Repetitive sequences in the kiwifruit genomes. Table S3. Summary of structure variations among Actinidia species. Table S4. The 222 unique SVs only identified from *A. latifolia* and *A. eriantha*. Table S5. Genes associated with SVs identified in Table S4. Table S6. Promoter regions associated with SVs identified in Table S4. Table S7. SV associated genes correlated with the content of vitamin C. Table S8. The 848 unique SVs only identified from *A. polygama* and *A. macrosperma*. Table S9. Genes associated with SVs identified in Table S8. Table S10. Promoter regions associated with SVs identified in Table S8. Table S11. GO enrichment of gene families expanded in *A. latifolia* and *A. eriantha*. Table S12. GO enrichment of gene families contracted in *A. polygama* and *A. macrosperma*.

Acknowledgements

The authors thank professor Fengwang Ma from the Northwest A&F University for providing help during sample collection.

Authors' contributions

X.S. designed the project. L.H. collected the samples for sequencing. X.L1. (Xiaolong Li), C.Z., X.S. analyzed the data. X.L2. (Xinyi Li), M.G., C.X., X.H., J.G. performed root transformation and VC measurement. J.F., Y.Z. built the genome database. X.L1. drafted the manuscript. X.S. revised the manuscript. All authors read and commented on the final manuscript.

Funding

This study was supported by the Zhejiang Provincial Natural Science Foundation of China (LR23C150001; LQ24C150003; Q24C150007), the National Natural Science Foundation of China (NSFC; 32102318), and the NSFC Excellent Young Scientists Fund (overseas; to X.S.).

Availability of data and materials

Raw reads for genome sequencing and the assemblies have been deposited in the National Genomics Data Center (NGDC; <https://ngdc.cncb.ac.cn>) under the BioProject no. PRJCA026928 [63]. Genome assemblies and annotations are also available through our customized kiwifruit database: <http://actinidiabase.moilab.net> [64]. Other Actinidia genomes and annotation files used for comparative genome are available under the DOI 10.1038/s41438-019-0202-y (*A. chinensis*) [65], DOI 10.1038/s41597-023-02006-4 (*A. deliciosa*) [66], DOI 10.1016/j.molp.2022.12.022 (*A. latifolia*) [10], and DOI 10.1093/gigascience/giz027 (*A. eriantha*) [67].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have read and given their approval to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Author details

¹Collaborative Innovation Center for Efficient and Green Production of Agriculture in Mountainous Areas of Zhejiang Province, College of Horticulture Science, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China. ²Key Laboratory of Quality and Safety Control for Subtropical Fruit and Vegetable, Ministry of Agriculture and Rural Affairs, Zhejiang A&F University, Hangzhou, Zhejiang 311300, China. ³State Key Laboratory of Rice Biology, Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Biotechnology, Zhejiang University, Hangzhou, Zhejiang 310058, China. ⁴Bioinformatics Center, Beijing University of Agriculture, Beijing 102206, China.

Received: 28 July 2024 Accepted: 3 September 2024
Published online: 11 September 2024

References

- Ferguson A. Kiwifruit: a botanical review. In: Janick J, editor. Horticultural reviews. Hoboken: Wiley; 2011. p. 1–64.
- Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, et al. Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun*. 2013;4:2640.
- Morton J, Dowling C. Fruits of warm climates. Miami: Winterville: J.F. Morton; Distributed by Creative Resources Systems; 1987.
- Smith G, Walton E. Kiwifruit. In: Erez A, editor. Temperate fruit crops in warm climates. Dordrecht: Springer Netherlands; 2000. p. 367–79.
- Liu X, Wu R, Bulley S, Zhong C, Li D. Kiwifruit MYB51-like and GBF3 transcription factors influence L-ascorbic acid biosynthesis by activating transcription of GDP-L-galactose phosphorylase 3. *New Phytol*. 2022;234:1782–800.
- Liu Y, Li D, Zhang Q, Song C, Zhong C, Zhang X, et al. Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification. *New Phytol*. 2017;215:877–90.
- Pilkington S, Crowhurst R, Hilario E, Nardoza S, Fraser L, Peng Y, et al. A manually annotated *Actinidia chinensis* var. *chinensis* (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC Genomics*. 2018;19:257.
- Wu H, Ma T, Kang M, Ai F, Zhang J, Dong G, Liu J. A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic Res*. 2019;6:117.
- Yue J, Chen Q, Wang Y, Zhang L, Ye C, Wang X, et al. Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit *Actinidia chinensis*. *Hort Res*. 2023;10:uhac264.
- Han X, Zhang Y, Zhang Q, Ma N, Liu X, Tao W, et al. Two haplotype-resolved, gap-free genome assemblies for *Actinidia latifolia* and *Actinidia chinensis* shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. *Mol Plant*. 2023;16:452–70.
- Akagi T, Varkonyi-Gasic E, Shirasawa K, Catanach A, Henry I, Merten D, et al. Recurrent neo-sex chromosome evolution in kiwifruit. *Nat Plants*. 2023;9:393–402.
- Wang Y, Li P, Zhu Y, Zhang F, Zhang S, He Y, et al. Graph-based pangenome of *Actinidia chinensis* reveals structural variations mediating fruit greening. *Adv Sci*. 2024;11:2400322.
- Bulley S, Rassam M, Hoser D, Otto W, Schünemann N, Wright M, et al. Gene expression studies in kiwifruit and gene over-expression in *Arabidopsis* indicates that GDP-L-galactose guanylyltransferase is a major control point of vitamin C biosynthesis. *J Exp Bot*. 2009;60:765–78.
- Xiao L, Xiao Y, Wang Z, Tan H, Tang K, Zhang L. Metabolic engineering of vitamin C production in *Arabidopsis*. *Biotechnol Bioproc E*. 2015;20:677–84.
- Zhou Y, Tao QC, Wang ZN, Fan R, Li Y, Sun XF, et al. Engineering ascorbic acid biosynthetic pathway in *Arabidopsis* leaves by single and double gene transformation. *Biologia Plant*. 2012;56:451–7.
- Shu P, Zhang Z, Wu Y, Chen Y, Li K, Deng H, et al. A comprehensive metabolic map reveals major quality regulations in red-flesh kiwifruit (*Actinidia chinensis*). *New Phytol*. 2023;238:2064–79.
- Simão F, Waterhouse R, Ioannidis P, Kriventseva E, Zdobnov E. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
- Kumar S, Stecher G, Suleski M, Hedges S. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–9.
- Xia H, Deng H, Li M, Xie Y, Lin L, Zhang H, et al. Chromosome-scale genome assembly of a natural diploid kiwifruit (*Actinidia chinensis* var. *deliciosa*). *Sci Data*. 2023;10:92.
- Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6.
- Tang W, Sun X, Yue J, Tang X, Jiao C, Yang Y, et al. Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin interaction mapping. *GigaScience*. 2019;8:gjz027.
- Ali M, Liu M, Wang Z, Li S, Jiang T, Zheng X. Pre-harvest spraying of oxalic acid improves postharvest quality associated with increase in ascorbic acid and regulation of ethanol fermentation in kiwifruit cv. Bruno during storage. *J Integr Agric*. 2019;18:2514–20.
- Su W, Ren Y, Wang D, Su Y, Feng J, Zhang C, et al. The alcohol dehydrogenase gene family in sugarcane and its involvement in cold stress regulation. *BMC Genomics*. 2020;21:521.
- Xing C, Liu Y, Zhao L, Zhang S, Huang X. A novel MYB transcription factor regulates ascorbic acid synthesis and affects cold tolerance: PbrMYB5 functions in cold tolerance. *Plant Cell Environ*. 2019;42:832–45.
- Zhang H, Wu K, Wang Y, Peng Y, Hu F, Wen L, et al. A WUSCHEL-like homeobox gene, OsWOX3B responses to NUDA/GL-1 locus in rice. *Rice*. 2012;5:30.
- Rahmani R, Decap D, Fostier J, Marchal K. BLSSpeller to discover novel regulatory motifs in maize. *DNA Res*. 2022;29:dsac029.
- Hu X, Li T, Xu C, Xu Y, You C, Li X, Gong J, et al. Identification and characterization of genes related to m⁶A modification in kiwifruit using RNA-seq and ATAC-seq. *Hort Plant J*. 2024;10(5):1121–33. <https://doi.org/10.1016/j.hpj.2024.02.001>.
- O'Malley R, Huang S, Song L, Lewsey M, Bartlett A, Nery J, Galli M, et al. Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell*. 2016;165:1280–92.
- Zhao M, Morohashi K, Hatlestad G, Grotewold E, Lloyd A. The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development*. 2008;135:1991–9.
- Hancock R, Viola R. Biosynthesis and catabolism of L-ascorbic acid in plants. *Crit Rev Plant Sci*. 2005;24:167–88.
- Sawake S, Tajima N, Mortimer J, Lao J, Ishikawa T, Yu X, et al. KON-JAC1 and 2 are key factors for GDP-mannose generation and affect L-ascorbic acid and glucomannan biosynthesis in *Arabidopsis*. *Plant Cell*. 2015;27:3397–409.
- Liu X, Xie X, Zhong C, Li D. Comparative transcriptome analysis revealed the key genes regulating ascorbic acid synthesis in *Actinidia*. *Int J Mol Sci*. 2021;22:12894.
- Gong J, Chen Y, Xu Y, Gu M, Ma H, Hu X, Li X, et al. Tracking organelle activities through efficient and stable root genetic transformation system in woody plants. *Hort Res*. 2024;11:uhad262.
- Sun X, Jiao C, Schwanning H, Chao C, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020;52:1423–32.
- Li X, Li X, Li J, Soejarto D. Progress in the phylogeny and taxonomy of *Actinidia* during the past decade. *Acta Hort*. 2011;913:71–6.
- Hu J, Wang Z, Sun Z, Hu B, Ayoola A, Liang F, Li J, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol*. 2024;25:107.
- Chen Y, Nie F, Xie S, Zheng Y, Dai Q, Bray T, Wang Y, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun*. 2021;12:60.
- Roach M, Schmidt S, Borneman A. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*. 2018;19:1–10.
- Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*. 2020;36:2253–5.
- Chakraborty M, Baldwin-Brown J, Long A, Emerson J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res*. 2016;44:e147.
- Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. 2021;23:258.
- Ou S, Su W, Liao Y, Chougule K, Agda J, Hellinga A, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol*. 2019;20:275.
- Brüna T, Hoff K, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*. 2021;3:lqaa108.
- Gabriel L, Hoff K, Brüna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*. 2021;22:566.
- Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.

46. Emms D, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238.
47. De Bie T, Cristianini N, Demuth J, Hahn M. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics.* 2006;22:1269–71.
48. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
49. Wang Y, Tang H, DeBarry J, Tan X, Li J, Wang X, et al. MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49.
50. Marçais G, Delcher A, Phillippy A, Coston R, Salzberg S, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol.* 2018;14:e1005944.
51. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell.* 2021;184:3542–3558.e16.
52. Goel M, Sun H, Jiao W, Schneeberger K. SyRl: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 2019;20:277.
53. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet.* 2014;46:919–25.
54. Liang P, Saqib H, Zhang X, Zhang L, Tang H. Single-base resolution map of evolutionary constraints and annotation of conserved elements across major grass genomes. *Genome Biol Evol.* 2018;10:473–88.
55. Kielbasa S, Wan R, Sato K, Horton P, Frith M. Adaptive seeds tame genomic sequence comparison. *Genome Res.* 2011;21:487–93.
56. Frith M. A new repeat-masking method enables specific detection of homologous sequences. *Nucleic Acids Res.* 2011;39:e23–e23.
57. Frith M, Kawaguchi R. Split-alignment of genomes finds orthologies more accurately. *Genome Biol.* 2015;16:106.
58. Kent W, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci USA.* 2003;100:11484–9.
59. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90.
60. Kim D, Paggi J, Park C, Bennett C, Salzberg S. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15.
61. Liao Y, Smyth G, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30.
62. Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
63. Li X, Huo L, Li X, Zhang C, Gu M, Fan J, et al. Genomes of diverse Actinidia species provide insights into *cis*-regulatory motifs and genes associated with critical traits. 2024. NGDC accession <https://ngdc.cncb.ac.cn/search/all?q=PRJCA026928>.
64. Li X, Huo L, Li X, Zhang C, Gu M, Fan J, et al. Genomes of diverse Actinidia species provide insights into *cis*-regulatory motifs and genes associated with critical traits. 2024. Kiwifruit database <http://actinidiabase.moilab.net>.
65. Wu H, Ma T, Kang M, Ai F, Zhang J, Dong G, Liu J. A high-quality *Actinidia chinensis* (kiwifruit) genome. GenBank; 2019. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA549770>.
66. Xia H, Deng H, Li M, Xie Y, Lin L, Zhang H, et al. Chromosome-scale genome assembly of a natural diploid kiwifruit (*Actinidia chinensis* var. *deliciosa*). Genbank; 2023. <https://www.ncbi.nlm.nih.gov/datasets/genome/?bioproject=PRJNA850524>.
67. Tang W, Sun X, Yue J, Tang X, Jiao C, Yang Y, et al. Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin interaction mapping. Genbank; 2019. <https://www.ncbi.nlm.nih.gov/nucleotide/QOV500000000>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.