



OPEN

DATA DESCRIPTOR

Chromosome-scale genome assembly of a natural diploid kiwifruit (*Actinidia chinensis* var. *deliciosa*)

Hui Xia^{1,3}, Honghong Deng^{1,3}, Mingzhang Li², Yue Xie², Lijin Lin¹, Huifen Zhang¹, Xian Luo¹, Xiulan Lv¹, Jin Wang¹ & Dong Liang¹✉

The most commercialized kiwifruit, *Actinidia chinensis* var. *deliciosa* (*Acd*), is an allohexaploid ($2n = 6x = 174$), making high-quality assemblage genome challenging. We previously discovered a rare naturally occurring diploid *Acd* plant. Here, chromosome-level *de novo* genome assembly for this diploid *Acd* was reported, reaching approximately 621.98 Mb in length with contig and scaffold N50 values of 10.08 and 21.09 Mb, respectively, 99.66% of the bases anchored to 29 pseudochromosomes, and 38,990 protein-coding genes and 42.29% repetitive elements annotated. The divergence time of *A. chinensis* cv. 'Red5' and 'Hongyang' (11.1–27.7 mya) was more recent compared with the divergence time of them and *Acd* (19.9–41.2 mya), with the divergence time of *A. eriantha* cv. 'White' being the earliest (22.9–45.7 mya) among that of the four *Actinidia* species. The 4DTV distance distribution highlighted three recent whole-genome duplication events in *Acd*. This is the first high-quality diploid *Acd* genome, which lays an important foundation for not only kiwifruit functional genomics studies but also further elucidating genome evolution of allohexaploid *Acd*.

Background & Summary

Kiwifruit, belonging to the genus *Actinidia* Lindl. and the Actinidiaceae family (Ericales), is an economically and nutritionally important fruit crop worldwide¹. China is the world's largest kiwifruit producer, reaching an annual production of approximately 2.23 million tons, representing slightly more than half of the global annual production (4.34 million tons)². As consumers become increasingly health-conscious, kiwifruit has attracted special attention because it is a nutrient-rich fruit and an ample source of vitamin C, dietary fibers, carotenoids, chlorophylls, and phenols^{3,4}. Its nutritional qualities have made kiwifruit known as “the king of the fruits” for centuries⁵.

As sequencing technology and associated bioinformatics tools have advanced, kiwifruit genomic studies have followed suit. To date, four kiwifruit genomes have been published⁶. In 2013, the first draft assembly of *A. chinensis* var. *chinensis* cv. 'Hongyang', a diploid heterozygous kiwifruit, was assembled using only short Illumina reads⁵. The genome of *A. chinensis* var. *chinensis* cv. 'Red 5' was assembled using short Roche 454 reads⁷. The advent of third-generation sequencing (TGS) technology, such as the Pacific Biosciences (PacBio, Menlo Park, CA, USA) long-read single-molecule real-time (SMRT) sequencing technology, has led to the chromosome-scale genome assembly of *A. eriantha* cv. 'White'⁸ and an improved genome assembly of 'Hongyang', as evidenced by the increased contig and scaffold N50 values⁹.

PacBio is a 'sequencing by synthesis' platform similar as to Illumina sequencing, whereas the Oxford Nanopore Technologies (ONT) sequencing, (Oxford Nanopore Technologies, Oxford, UK) is another important TGS method that adopts a nanopore-based single-molecule sequencing strategy. TGS enables the sequencing of very long fragments, up to 30 to 50 kb or even longer. This advantage of TGS has made it a very attractive and suitable option for the high-quality assembly of plant genomes, even if the genomes are extremely large and contain many highly repetitive DNA stretches¹⁰.

¹Institute of Pomology and Olericulture, College of Horticulture, Sichuan Agricultural University, Chengdu, 611130, China. ²Key Laboratory of Breeding and Utilization of Kiwifruit in Sichuan Province, Sichuan Provincial Academy of Natural Resource Sciences, Chengdu, China. ³These authors contributed equally: Hui Xia, Honghong Deng. ✉e-mail: liangeast@sicau.edu.cn



Fig. 1 Fruit phenotype of the natural diploid *Actinidia chinensis* var. *deliciosa* resource.

In the genus *Actinidia*, the basic chromosome number is 29, and the ploidy ranges from diploid to hexaploid, and may also be octoploid, decaploid, or dodecaploid¹¹. Many species within *Actinidia* bear edible fruits, but four species, including *A. chinensis* Planch. var. *chinensis*, *A. chinensis* var. *deliciosa* (A. Chev.) A. Chev., *A. arguta* (Sieb. et Zucc.) Planch. ex Miq., and *A. eriantha* Benth., are currently cultivated extensively and economically important¹. An allohexaploid ($2n = 6x = 174$) *A. chinensis* cv. ‘Hayward’ has been the most popular worldwide and dominated the international market for many years¹². Green-fleshed kiwifruit like ‘Hayward’ is by now the most commercialized. Despite the importance of *A. chinensis* var. *deliciosa* (*Acd*) in the kiwifruit industry, no genomic information is available for this species. Generating a high-quality *de novo* *Acd* genome has been challenging because of its high ploidy and large proportion of repetitive DNA sequences.

Over the past few years, most genomic studies have used the diploid *A. chinensis* var. *chinensis* cv. ‘Hongyang’. Most previous studies on *Acd* have used the hexaploid ‘Harward’, but it only represents one pistillate cultivar from one variety (var. *deliciosa*) belong to one species (*A. chinensis*). The ploidy variation and morphological characteristics of 600 wild populations of *A. chinensis* var. *chinensis* and *Acd* were examined by Li *et al.* (2010), and they found that the majority of *Acd* were classified as tetraploid and hexaploid, with two exceptions being pentaploid and one octoploid¹³.

In our kiwifruit breeding program, an unexpected and naturally occurring diploid *Acd* plant was identified and obtained in our previous study (Fig. 1). This is an ideal natural material for unveiling the genomic structure of *Acd*. The benefits of diploid kiwifruit cultivars have also materialized in elucidating the molecular mechanisms underlying multiple development aspects, such as flowering, fruit ripening, and color and flavor development^{14–17}. The wild diploid *Acd* plant resource was reported for the first time, which is a rare resource among the natural complex ploidy races in *Acd*. Additionally, wild species may have many important characteristics that can be used to improve the kiwifruit cultivation. Deciphering genomic information is the first step toward maximizing the usefulness of this valuable resource.

In the present work, a high-quality chromosome-scale genome assembly of the natural diploid *Acd* was constructed by combining ONT sequencing, Illumina sequencing, and chromosome conformation capture (Hi-C) anchoring, which resulted in a genome approximately 621.98 Mb in length with contig and scaffold N50 values of 10.08 and 21.09 Mb, respectively (Table 1). Contigs were scaffolded into 29 superscaffolds, accounting for 99.66% of the total genome size. As shown in the Hi-C heatmap (Fig. 2), the 29 superscaffolds in the *Acd* genome could be distinguished and perfectly represented by 29 chromosomes. The genome assembly presented here is less fragmented and more complete than the four previously published genome assemblies of kiwifruit (Table 1).

Comparative genomic analyses were also conducted on the *Acd* genome and other genomes of representative plant species to explore gene and genome evolution. First-ever high-quality chromosome-scale genome for *Acd* was reported in this study. The availability of the genome information provides a critical foundation for not only the phylogeny, genetic diversity, functional genomics, and genomics-assisted breeding studies of kiwifruit but also elucidating genome evolution of allohexaploid *Acd* in future.

Methods

Plant materials. Young tender leaf samples were randomly collected from the natural diploid *Acd* plant grown at the experimental base of the Sichuan Provincial Academy of Natural Resource Sciences, Deyang, China (N31°30′, E104°23′). Samples were immediately frozen in liquid nitrogen for Illumina and ONT sequencing. Ploidy was pre-confirmed using flow cytometry. Five types of tissues (leaves, fruit, buds, roots, and branches)

Parameter	<i>Actinidia deliciosa</i>	<i>Actinidia eriantha</i>	<i>Actinidia chinensis</i> (v3.0)	<i>Actinidia chinensis</i> (v2.0)	<i>Actinidia chinensis</i> (v1.0)
Total contig length (Mb)	621.98	690.4	653.86	—	604.22
Total contig No.	132	4,076	2,366	—	26,721
Contig N50 (kb)	10,083.34	539.2	1,430	—	58.86
Contig N90 (kb)	3,107.32	50.7	127.9	—	11.6
Longest contig length (Mb)	26.88	3.26	7.8	—	0.42
Total scaffold length (Mb)	621.99	690.6	653.93	548	616.11
Total scaffold No.	41	1,735	1,720	3,887	7,698
Scaffold N50 (kb)	21,085.33	23,583.90	20,000.00	623.8	646.79
Scaffold N90 (kb)	17,895.51	20,112.10	264.7	—	122.7
Longest scaffold length (Mb)	28.56	28.6	27.3	4.43	3.41
Anchored to chromosome (Mb/%)	619.87/99.66	682.4/98.84	640.75/97.98		

Table 1. Comparison of the assemblies of *Actinidia chinensis* var. *deliciosa* and the other four published kiwifruit genomes.

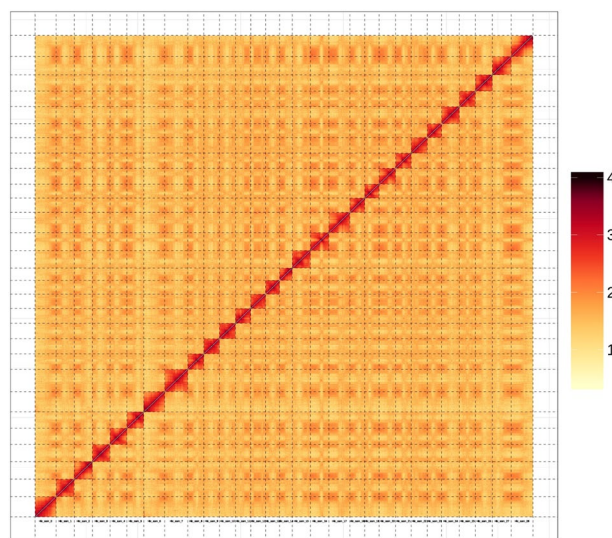


Fig. 2 Hi-C interaction heatmap for diploid *Actinidia chinensis* var. *deliciosa* genome. The map shows scaffolded and independently assembled chromosomes at high resolution.

were collected and mixed for RNA extraction. A combination of Illumina paired-end (PE), ONT, and Hi-C sequencing was performed by Novogene Co. Ltd. (Beijing, China).

Illumina sequencing and preliminary genome survey. In order to generate PE genome sequencing libraries with a 350-bp insert size, an Illumina TruSeq[®] Nano DNA library preparation kit (Illumina, San Diego, CA, USA) was used. The PE libraries were sequenced on an Illumina HiSeq X10 platform using the HiSeq X Ten reagent kit v2.5 (Illumina) for 2 × 150 cycles. Contaminants, low-quality reads, and Illumina adapters were trimmed using “fastp”¹⁸ prior to downstream genome size estimation. The clean PE reads were subjected to a K-mer counting algorithm using Jellyfish software v.2.2.10¹⁹. The abundance of 17-K-mer was determined using Kmerfreq v5.0²⁰ and provided an estimated genome size of 642.96 Mb with a 0.58% heterozygosity level and a repetitive DNA sequence rate of 49.73%. SOAPdenovo software v.2.0.4²⁰ was used for initial assembly. The results were then assembled into scaffold using a K-mer parameter of 41, based on the project experience of Novogene, which yielded 2,818 contigs and a scaffold N50 value of 4,791 bp.

RNA extraction and purification were performed using the Qiagen RNeasy Plant Mini Kit (Qiagen, CA, USA), following the manufacturer’s instructions. Equal amounts of RNA extracted from different tissues, including leaves, fruits, buds, roots, and branches, were mixed for RNA-Seq.

ONT library preparation, sequencing, and assembly. Genomic DNA (10 μg) was mechanically sheared into 10–50 kb fragments using a g-TUBE device (Covaris, Inc., MA, USA). Size selection was performed using the BluePippin size selection system (Sage Science, Inc., MA, USA). DNA end repair was performed using NEBNext End Repair/dA-Tailing Module (New England Biolabs, MA, UK). ONT PromethION sequencing libraries were prepared using the SQK-LSK109 ligation sequencing kit (Oxford Nanopore Technologies) as per the manufacturer’s standard protocol.

Repeat family	De novo + Repbase		TE proteins		Combined Tes	
	Length (bp)	Proportion in genome (%)	Length (bp)	Proportion in genome (%)	Length (bp)	Proportion in genome (%)
DNA transposons	3,819,086	0.61	126,403	0.02	3,881,658	0.62
LINE	1,865,614	0.3	140,320	0.02	1,915,779	0.31
SINE	10,314	0	0	0	10,314	0
LTR	#####	34.87	46,054,847	7.4	#####	35.55
Unknown	30,012,173	4.83	0	0	30,012,173	4.83
Total (TRF not included)	#####	40.06	46,321,434	7.45	#####	40.58

Table 2. Repetitive elements and their proportions in diploid *Actinidia chinensis* var. *deliciosa* genome.

After the quality control, NextDenovo software v.2.3.0 (<https://github.com/Nextomics/NextDenovo>) was applied to correct the error rate of Nanopore long reads using the Illumina short reads. The preliminary assembly was further polished by Illumina PE reads using the NextPolish software v.1.2.4. The redundant contigs led by heterozygosity were identified and removed using Purge Haplotigs. After filtering low-quality reads, 80.32 Gb of Illumina short-read and 79.0 Gb of ONT long-read data were obtained, representing approximately 124.92- and 122.87-fold coverage of the estimated genome (642.96 Mb), respectively.

Hi-C assisted scaffolding. Hi-C library preparation and sequencing followed the protocol of Belton *et al.* (2012)²¹ with minor modifications. The chromatin was cross-linked with a 4% formaldehyde solution. The cross-linked nuclei were digested using restriction enzymes targeting GATC and GANTC restriction sites. A biotin-labeled nucleotide was incorporated at the digested ends, and the ends were ligated. Ligation products were purified and prepared to an appropriate size (150 bp) for Illumina short-read sequencing. The Hi-C reads were linked into pseudochromosomes using ALLHiC software v.0.9.14. After ALLHiC scaffolding, Juicebox v1.13.01 was used to manually correct large-scale inversions and translocations to obtain the final pseudochromosomes.

Repetitive element identification. An alignment of homologous sequence and a *de novo* prediction method were combined to identify repetitive elements. The former adopted the RepeatMasker v.4.1.0 and Repbase library (<http://www.girinst.org/replib/>) to identify known transposable elements (TEs), which were then aligned with the genome sequences from a TE protein database, RepeatProteinMask (<http://www.repeatmasker.org/>). The latter applied the LTR_FINDER v.1.0.6 (http://tlife.fudan.edu.cn/ltr_finder/), RepeatScout v.1.0.5 (<http://www.repeatmasker.org/>), and RepeatModeler v.2.0.1 (<http://www.repeatmasker.org/RepeatModeler.html>) to construct a *de novo* repeat library, and then adopted the RepeatMasker v.4.1.0 to predict the repetitive element family.

In total, 42.29% of the assembled sequences were identified as repetitive DNA sequences. The most abundant repetitive DNA sequences were retrotransposons (35.85%), of which the long terminal repeat (LTR) retrotransposons accounted for 35.55% with Gypsy (13.99%) and Copia (13.23%) elements being dominant. The DNA transposons represented 0.62% of the genome (Table 2).

Protein-coding genes prediction. The protein-coding genes were annotated by incorporating transcriptome data, homology-based searches, and *de novo* predictions, resulting in 38,990 genes. In the homology-based prediction, we used TblastN v.2.2.26 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and GeneWise software v.2.4.1. Augustus v.3.2.3²², Geneid v.1.4²³, Genescan v.1.0²⁴, GlimmerHMM v.3.04²⁵, and SNAP v.2013.11.29²⁶ were used for *de novo* prediction. *De novo* transcriptome assembly was performed using Trinity v.2.1.1²⁷. Hisat v.2.0.4²⁸/TopHat v.2.0.11²⁹ was used to align RNA-Seq reads to the genome FASTA to identify the exons and splice sites. Stringtie v.1.3.3³⁰/Cufflinks v.2.2.1 was used to assemble transcripts based on the alignment results. To generate a consensus gene set, we used EvidenceModeler (EVM) v.1.1.1 and the Program to Assemble Spliced Alignments³¹ to merge genes predicted in the three respective annotation files.

The average transcript length and coding sequence size were 5,457.64 and 1,178.37 bp, respectively, with a mean of 5.28 exons per gene. The average exon and intron lengths were 223 and 1,000 bp, respectively. The number of annotated genes in *Acd* was close to that in the *A. chinensis* v3.0 genome, which has 40,466 genes (Table 3).

Gene function annotation. Functional annotation of the predicted genes was performed by aligning them to SwissProt, NCBI non-redundant protein (Nr), Kyoto Encyclopedia of Genes and Genomes (KEGG), InterPro (<https://www.ebi.ac.uk/interpro/>), Gene Ontology (GO), and protein families (Pfam). A search of publicly available databases (ProDom³², PRINTS³³, Pfam³⁴, SMART³⁵, PANTHER³⁶, and PROSITE³⁷) was conducted using InterProScan v.5.27.66 to annotate the motifs and domains. A total of 38,079 genes (97.70% of the predicted protein-coding genes) were annotated using the above multiple databases. Specifically, approximately 78.00%, 96.60%, 75.30%, 94.70%, 59.10%, and 74.10% were annotated in SwissProt, Nr, KEGG, InterPro, GO, and Pfam, respectively.

Non-coding RNA annotation. Prediction of transfer RNAs (tRNAs) was conducted using tRNAscan-SE v.1.3.1 (<http://lowelab.ucsc.edu/tRNAscan-SE/>). We used BLAST to predict rRNAs based on the conservation of plant rRNA sequences. Other non-coding RNAs (ncRNAs), including microRNAs (miRNAs) and small nuclear RNAs (snRNAs), were identified using Infernal software v.1.1.2 (<http://infernal.janelia.org/>) by searching against the Rfam v14.1 database. The predicted non-coding genes included 5,626 ncRNAs, including 967 miRNAs, 705 tRNAs, 1,042 rRNAs, and 2,912 snRNAs in the *Acd* genome.

Species	Number	Average transcript length(bp)	Average CDS length(bp)	Average exons per gene	Average exon length(bp)	Average intron length(bp)
<i>Actinidia chinensis</i> var. <i>deliciosa</i>	38,990	5,457.64	1,178.37	5.28	223.26	1,000.28
<i>Actiniida eriantha</i>	42,988	4,809.50	1,006.73	5.07	198.72	935.22
<i>Actinidia chinensis</i> (v3.0)	40,466	4,900.83	1,027.86	5.3	193.91	900.55
<i>Actinidia chinensis</i> (v2.0)	33,124	5,140.63	1,278.43	5.47	233.78	864.3

Table 3. Comparison of the gene structure of the diploid *Actinidia chinensis* var. *deliciosa* and other published kiwifruit genomes.

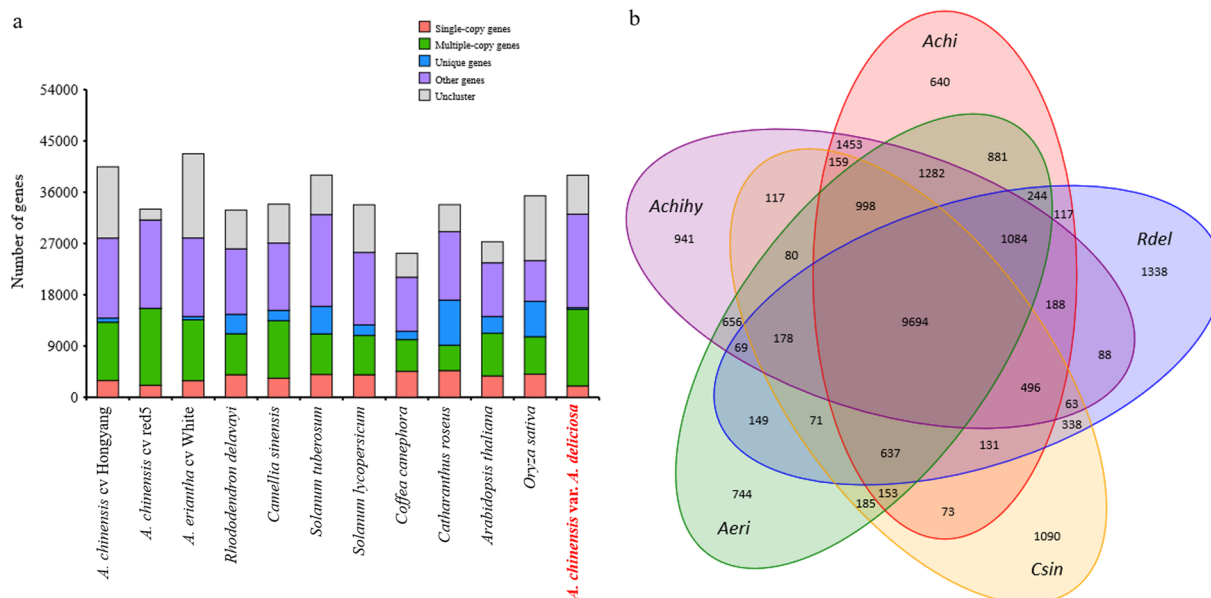


Fig. 3 Gene distribution in *Actinidia chinensis* var. *deliciosa* and 11 other representative species (a) and the common and unique gene families in *Actinidia chinensis* var. *deliciosa* and four other species (b). The single-copy genes and multiple-copy genes are presented in all the 12 plant species. The unique genes are only presented in one of these 12 plant species and the other genes are those presented in two to eleven plant species.

Comparative genomics analysis. Orthologous relationships between the genes of the currently assembled genome of *Acd* and other 11 representative plant species, including *A. chinensis* cv. ‘Hongyang’, *A. chinensis* cv. ‘Red5’, *A. eriantha* cv. ‘White’, *Rhododendron delavayi*, *Camellia sinensis*, *Solanum tuberosum*, *Solanum lycopersicum*, *Coffea canephora*, *Catharanthus roseus*, *Arabidopsis thaliana*, and *Oryza sativa*, were analyzed through all-against-all protein sequence similarity searches using OrthoMCL software v.2.0.9 (<http://orthomcl.org/orthomcl/>). For each orthologous cluster, we retained only the longest predicted transcript per locus. Clustering of these 12 plants’ protein-coding sequences yielded a total of 32,884 gene families. There were 6,159 common gene families and 694 common single-copy gene families (Fig. 3a), which represent the evolutionarily conserved and ancestral gene families. In *Acd*, 816 genes assigned to 640 gene families were unique (Fig. 3b).

The alignment was generated using Muscle v3.8.31 (<http://www.drive5.com/muscle/>) and was further refined using Gblocks 0.91b (<http://molevol.cmima.csic.es/castresana/Gblocks.html>). Based on protein sequences from single-copy orthologous groups, a maximum likelihood phylogenetic tree was constructed using RAxML v.8.2.12 (<http://sco.h-its.org/exelixis/web/software/raxml/index.html>). *Acd* was placed as a sister clade to the cultivated *A. chinensis* species (Red 5 and Hongyang), indicating its closer relationship with the *A. chinensis* than *A. eriantha*. Additionally, phylogenetic results showed that *Actinidia* species were most closely related to *R. delavayi* and *C. sinensis*. The divergence time with 95% confidence intervals among these 12 plant species was calculated using MCMCtree. The calibration points for divergence time estimation were taken from the TimeTree website (<http://www.timetree.org/>). The divergence time was as follows: *R. delavayi* - *A. chinensis*, 52.0–96.0 million years ago (mya); *C. canephora* - *C. roseus*, 61.0–95.0 mya; *S. tuberosum* - *S. lycopersicum*, 5.23–9.40 mya; *S. tuberosum* - *A. thaliana*, 111.0–131.0 mya; and *C. canephora* - *O. sativa*, 115.0–308.0 mya. The divergence time of the *A. chinensis* cv. ‘Red 5’ and *A. chinensis* cv. ‘Hongyang’ (11.1–27.7 mya) was more recent compared with the divergence time of *A. chinensis* cv. ‘Red 5’ and *A. chinensis* cv. ‘Hongyang’ and *Acd* (19.9–41.2 mya), with the divergence time of *A. eriantha* cv. ‘White’ being the earliest 22.9–45.7 mya among that of the four *Actinidia* species.

The maximum likelihood tree was used as a starting point for estimating species divergence time using the MCMCtree program (<http://abacus.gene.ucl.ac.uk/software/paml.html>), which was incorporated into the

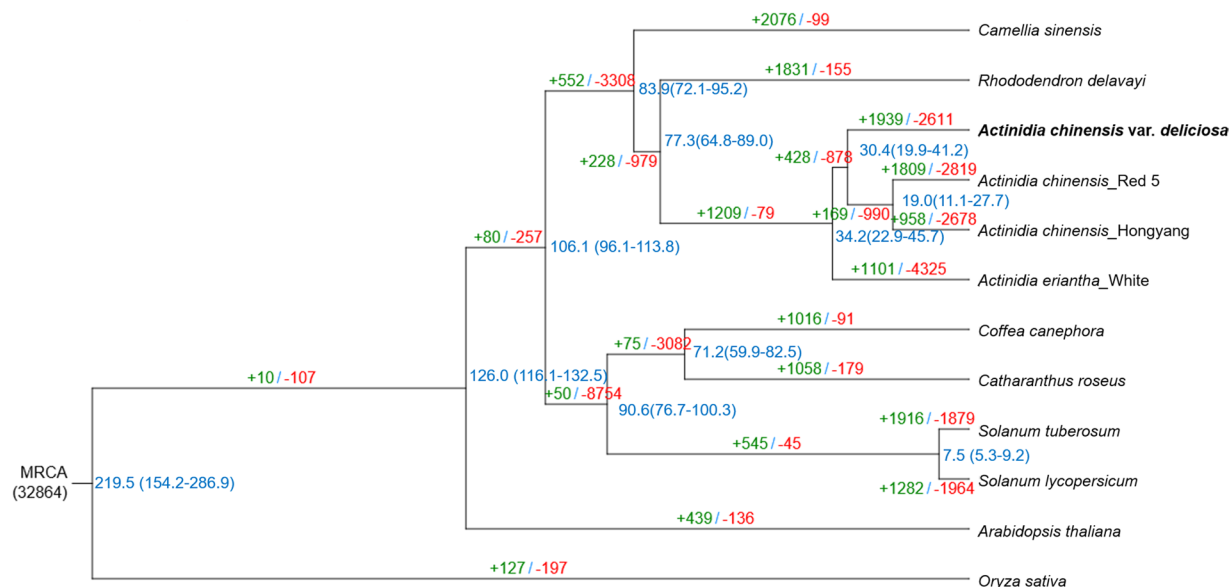


Fig. 4 Phylogenetic tree of the 12 species, and gene family expansion and contraction. Inferred divergence times are denoted at each node in blue. Gene family expansion and contraction are indicated in green and red, respectively.

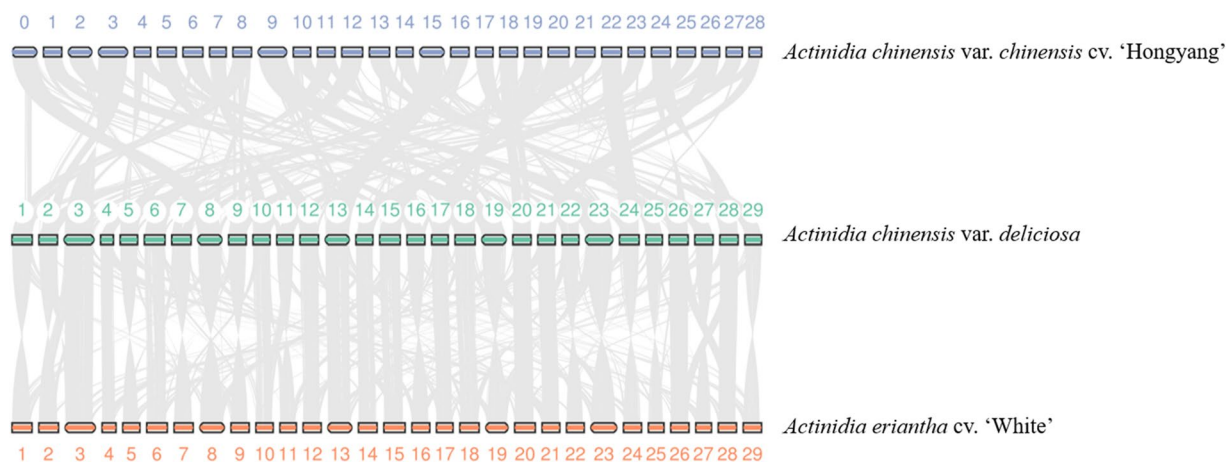


Fig. 5 Genome synteny among *Actinidia chinensis* var. *deliciosa*, *A. chinensis* cv. 'Hongyang' and *A. eriantha* cv. 'White'.

Phylogenetic Analysis using the Maximum Likelihood (PAML) software v4.9e. The expansion and contraction of the gene families were determined using Computation Analysis of gene Family Evolution (CAFE) v3.1 (<http://sourceforge.net/projects/cafehahnlab/>). In total, 1,939 and 2,611 gene families expanded and contracted in the *Acd* genome, respectively (Fig. 4).

Synteny and whole-genome duplication analysis. The homologous proteins between *Acd* and *A. chinensis* cv. 'Hongyang', *A. chinensis* cv. 'Red5', and *A. eriantha* cv. 'White' were identified using BLASTP. MCSanX³⁸ was used to assess the synteny between genomes of *Acd* and *A. chinensis* cv. 'Hongyang', *A. chinensis* cv. 'Red5', and *A. eriantha* cv. 'White', with at least five syntenic genes and no more than 15 gapped genes. Genome synteny analysis revealed a higher degree of synteny between the genomes of *Acd*, *A. chinensis* cv. 'Hongyang', and *A. eriantha* cv. 'White' (Fig. 5).

As a neutral genetic distance, four-fold degenerated sites (4DTv) has been widely used to analyze evolutionary divergence and estimate the relative timing of putative whole-genome duplication (WGD) events. The 4DTv value peaked at 0.05, 0.2, and 0.5, highlighting three recent WGD events in *Acd*. The speciation event that gave rise to the *Actinidia* species, represented by the pairwise 4DTv distribution of *A. eriantha* cv. 'White' against *R. delavayi*, occurred almost simultaneously with the duplication event found in *Actinidia* species (Fig. 6).

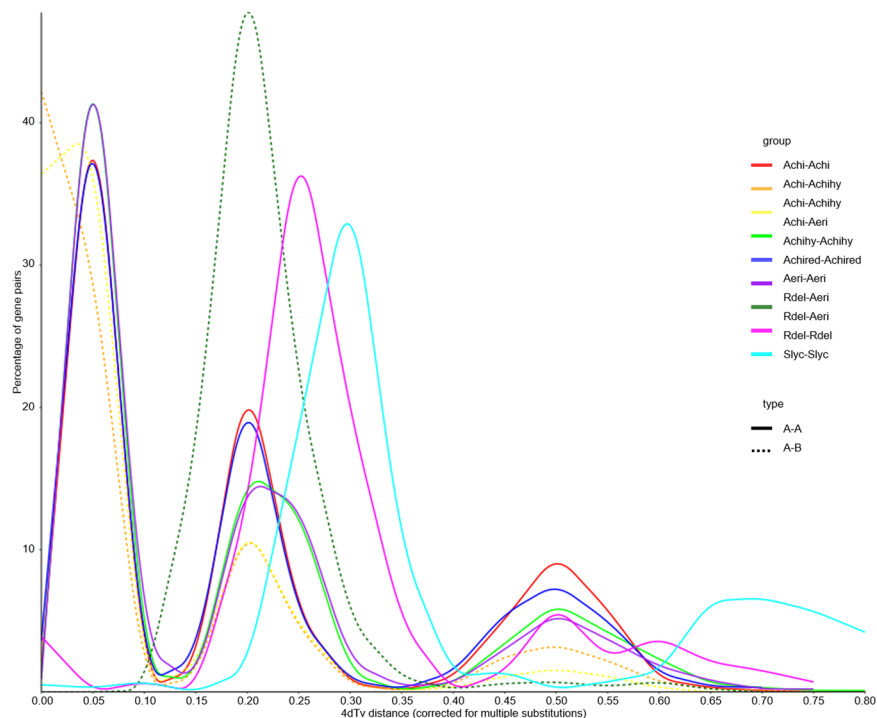


Fig. 6 Distribution of the 4DTV rates among the paralogs of the studied species. The dashed line shows the species divergence event, and the solid line shows the duplication event.

Data Records

The genome sequencing data have been deposited in the NCBI database under the project ID PRJNA850524³⁹ and SRX17463538⁴⁰. The genome estimation, assemble statistics based on 41-K-mer, assemble statistics based on Illumina and Nanopore sequencing data, BUSCO assessment, GEGMA assessment, Illumina reads coverage, single nucleotide polymorphisms, repetitive elements, gene function annotations, ncRNA prediction, gene families, gene expansion, and gene contraction have been deposited at the Figshare database with <https://doi.org/10.6084/m9.figshare.20279286.v2>⁴¹.

Technical Validation

The completeness of the final assembled genome was assessed using Benchmarking Universal Single-Copy Orthology (BUSCO) software v4.0.5 by searching plant-specific databases that contain a total of 1,440 orthologous single-copy genes. The results revealed the retrieval of 96.1% of the complete single-copy genes, of which 27.8% were duplicated. Only 0.7% of BUSCO genes were fragmented, and 3.2% were missing from the genome. The BUSCO results indicated high genome assembly completeness.

Core Eukaryotic Genes Mapping Approach (CEGMA, v2.0) employs 248 most highly conserved core eukaryotic genes (CEGs) to assess the extent of comprehensive gene coverage. The genome assembly showed a high completeness level, with 95.16% (236) of CEGs completely and partially covered, including 81.05% (201) of CEGs considered complete.

The filtered short Illumina reads were aligned back to the genome assembly using Burrows–Wheeler Aligner (BWA) software v.0.7.12-r1039 (<http://bio-bwa.sourceforge.net/>). Approximately 99.25% of the short reads mapped to the genome, and 98.72% were marked as properly paired. The ratios of heterozygous and homozygous single nucleotide polymorphisms (SNPs) were 0.422941% and 0.000588%, respectively, indicating that the assembly had high single-base-level accuracy.

Code availability

NextDenovo software v.2.3.0: read_cutoff = 1k, seed_cutoff = 26,457, sort_options = -m 20 g -t 10 -k 40, minimap2_options_raw = -x ava-ont -t 10, correction_options = -p 10, minimap2_options_cns = -x ava-ont -t 10 -k17 -w17, and nextgraph_options = -a 1.

NextPolish software v.1.2.4: sgs_options = -max_depth 200, lgs_options = -min_read_len 1k -max_read_len 100k -max_depth 100, and lgs_minimap2_options = -x map-ont.

ALLHiC software v.0.9.14: “-NonInformativeRatio 0”, “-minREs 50”, “-MaxLinkDensity 3”, “-shortest_150”, “-longest_800”, “-format_Sanger”, “-enz DpnII”, and “-CLUSTER 29”.

TblastN v.2.2.26: E-value $\leq 1e-5$.

Swiss-Prot: E-value $\leq 1e-5$.

Nr: E-value $\leq 1e-5$.

KEGG: E-value $\leq 1e-3$.

GO: E-value $\leq 1e-10$.

Pfam: E-value ≤ 0.01 .

BLAST to predict rRNAs: E-value $\leq 1e-5$, identify $\geq 85\%$, and match length ≥ 50 bp.

BLASTP: E-value $\leq 1e-5$.

Other commands and pipelines used in data processing were executed using their corresponding default parameters.

Received: 20 September 2022; Accepted: 7 February 2023;

Published online: 14 February 2023

References

- Ferguson, A. R. in *The Kiwifruit Genome, Compendium of Plant Genomes* (ed. Testolin, R., Huang, H. & Ferguson, A. R.) Ch.1 (Springer, 2016).
- FAOSTAT. Production Volume of Kiwis Worldwide in 2020, by Leading Country (in thousand metric tons) https://www.fao.org/faostat/en/#rankings/countries_by_commodity (2020).
- Liang, D. *et al.* Dynamic changes of phenolic compounds and their associated gene expression profiles occurring during fruit development and ripening of the Donghong kiwifruit. *J. Agric. Food Chem.* **68**, 11421–11433 (2020).
- Xia, H. *et al.* Changes in the carotenoids profile of two yellow-fleshed kiwifruit cultivars during storage. *Postharvest Biol. Technol.* **164**, 111162 (2020).
- Huang, S. *et al.* Draft genome of the kiwifruit *Actinidia chinensis*. *Nat. Commun.* **4**, 2640 (2013).
- Yue, J. *et al.* Kiwifruit genome database (KGD): A comprehensive resource for kiwifruit genomics. *Hortic. Res.* **7**, 117 (2020).
- Pilkington, S. M. *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* **19**, 257 (2018).
- Tang, W. *et al.* Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin interaction mapping. *Gigascience* **8**, giz027 (2019).
- Wu, H. *et al.* A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* **6**, 117 (2019).
- Dumschott, K., Schmidt, M. H. W., Chawla, H. S., Snowdon, R. & Usadel, B. Oxford Nanopore sequencing: New opportunities for plant genomics? *J. Exp. Bot.* **71**, 5313–5322 (2020).
- Ferguson, A. R. & Huang, H. in *The Kiwifruit Genome, Compendium of Plant Genomes* (ed. Testolin, R., Huang, H. & Ferguson, A. R.) Ch.5 (Springer, 2016).
- Salazar, J. *et al.* Transcriptome analysis and postharvest behavior of the kiwifruit '*Actinidia deliciosa*' reveal the role of ethylene-related phytohormones during fruit ripening. *Tree Genet. Genomes* **17**, 8 (2021).
- Li, D., Liu, Y., Zhong, C. & Huang, H. Morphological and cytotype variation of wild kiwifruit (*Actinidia chinensis* complex) along an altitudinal and longitudinal gradient in central-west China. *Bot. J. Linn. Soc.* **164**, 72–83 (2010).
- McAtee, P. A. *et al.* The hybrid non-ethylene and ethylene ripening response in kiwifruit (*Actinidia chinensis*) is associated with differential regulation of MADS-box transcription factors. *BMC Plant Biol.* **15**, 304 (2015).
- Ampomah-Dwamena, C. *et al.* A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* **221**, 309–325 (2019).
- Atkinson, R. G. *et al.* Dissecting the role of climacteric ethylene in kiwifruit (*Actinidia chinensis*) ripening using a 1-aminocyclopropane-1-carboxylic acid oxidase knockdown line. *J. Exp. Bot.* **62**, 3821–3835 (2011).
- Richardson, A. C. *et al.* Fruit development of the diploid kiwifruit, *Actinidia chinensis* 'Hort16A'. *BMC Plant Biol.* **11**, 182 (2011).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
- Luo, R. *et al.* SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
- Belton, J. M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
- Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
- Blanco, E., Parra, G. & Guigó, R. Using geneid to identify genes. *Curr. Protoc. Bioinforma.* **18**, 4.3.1–4.3.28 (2007).
- Aggarwal, G. & Ramaswamy, R. Ab initio gene identification: Prokaryote genome annotation with GeneScan and GLIMMER. *J. Biosci.* **27**, 7–14 (2002).
- Majoros, W. H., Perlea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
- Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
- Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
- Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- Perlea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
- Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
- Bru, C. *et al.* The ProDom database of protein domain families: More emphasis on 3D. *Nucleic Acids Res.* **33**, D212–D215 (2005).
- Attwood, T. K. The PRINTS database: a resource for identification of protein families. *Brief. Bioinform.* **3**, 252–263 (2002).
- Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: A web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234 (2000).
- Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–D288 (2005).
- Hulo, N. *et al.* The PROSITE database. *Nucleic Acids Res.* **34**, D227–D230 (2006).
- Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
- NCBI Assembly https://identifiers.org/insdc.gca:GCA_024454175.1 (2022).
- NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP395937> (2022).
- Figshare, <https://doi.org/10.6084/m9.figshare.20279286.v2> (2022).

Acknowledgements

This study was financially supported by grants of the Sichuan Science and Technology Program (2022YFH0049 and 2023YFH0003) and 'Xi Bu Zhi Guang' visiting scholar program to Hui Xia (supervisor: Jinbao Fang).

Author contributions

H.X. and D.L. conceived the idea, supervised the work, and revised the manuscript. M.L. and Y.X. prepared the plant materials. H.X., H.D., L.L., H.Z., X.L., X.L. and J.W. analyzed the data. H.D. wrote the original draft and revised the manuscript. H.X. and H.D. contributed equally to this work. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023