

GigaScience

Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin conformation capture --Manuscript Draft--

Manuscript Number:	GIGA-D-18-00282R2	
Full Title:	Chromosome-scale genome assembly of kiwifruit <i>Actinidia eriantha</i> with single-molecule sequencing and chromatin conformation capture	
Article Type:	Data Note	
Funding Information:	National Natural Science Foundation of China (31471157)	Dr. Yongsheng Liu
	National Science Foundation (IOS-1339287)	Dr. Zhangjun Fei
Abstract:	<p>Background: Kiwifruit (<i>Actinidia</i> spp.) is a dioecious plant with fruits containing abundant vitamin C and minerals. A handful of kiwifruit species have been domesticated, among which the <i>A. eriantha</i> is increasingly favored in breeding due to its superior commercial traits. Recently, elite cultivars from <i>A. eriantha</i> have been successfully selected and further studies on their biology and breeding potential require genomic information which is currently unavailable.</p> <p>Findings: Here, we assembled a chromosome-scale genome sequence of <i>A. eriantha</i> cv. White using single-molecular sequencing and chromatin conformation capture. The assembly has a total size of 690.6 Mb and an N50 of 21.7 Mb. Approximately 99% of the assembly were in 29 pseudomolecules corresponding to the 29 kiwifruit chromosomes. Forty-three percent of the <i>A. eriantha</i> genome are repetitive sequences, and the non-repetitive part encodes 42,850 protein-coding genes, of which 39,075 have homologues from other plant species or contain protein domains. The divergence time between <i>A. eriantha</i> and its close relative <i>A. chinensis</i> is estimated to be 3.3 million years, and after diversification, 1,740 and 1,345 gene families are expanded or contracted in <i>A. eriantha</i>, respectively.</p> <p>Conclusions: We generate a high-quality reference genome of kiwifruit <i>A. eriantha</i>. This chromosome-scale genome assembly is substantially better than two published kiwifruit assemblies from <i>A. chinensis</i> in terms of genome contiguity and completeness. The availability of <i>A. eriantha</i> genome provides a valuable resource for facilitating kiwifruit breeding and the studies of kiwifruit biology.</p>	
Corresponding Author:	Zhangjun Fei Boyce Thompson Institute for Plant Research Ithaca, NY UNITED STATES	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Boyce Thompson Institute for Plant Research	
Corresponding Author's Secondary Institution:		
First Author:	Zhangjun Fei	
First Author Secondary Information:		
Order of Authors:	Zhangjun Fei	
	Wei Tang	
	Xuepeng Sun	
	Junyang Yue	
	Xiaofeng Tang	
	Chen Jiao	
	Ying Yang	

	Xiangli Niu
	Min Miao
	Danfeng Zhang
	Shengxiong Huang
	Wei Shi
	Mingzhang Li
	Congbing Fang
	Yongsheng Liu
Order of Authors Secondary Information:	
Response to Reviewers:	Please check our response letter as a supplementary pdf file.
Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics	Yes
<p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	
Resources	Yes
<p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	

<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p>	<p>Yes</p>
---	------------

[Click here to view linked References](#)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 **Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule**
2 **sequencing and chromatin interaction mapping**

3
4 Wei Tang^{1,2,3*}, Xuepeng Sun^{4*}, Junyang Yue^{1,3*}, Xiaofeng Tang^{1,3}, Chen Jiao⁴, Ying Yang¹,
5 Xiangli Niu^{1,3}, Min Miao^{1,3}, Danfeng Zhang³, Shengxiong Huang³, Wei Shi³, Mingzhang Li⁵,
6 Congbing Fang¹, Zhangjun Fei^{4,6*}, Yongsheng Liu^{1,2,3*}

7
8 ¹School of Horticulture, Anhui Agricultural University, Hefei 230036, China

9 ²Ministry of Education Key Laboratory for Bio-resource and Eco-environment, College of Life
10 Science, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University,
11 Chengdu 610064, China

12 ³School of Food Science and Engineering, Hefei University of Technology, Hefei 230009, China

13 ⁴Boyce Thompson Institute, Cornell University, Ithaca NY 14853, USA

14 ⁵Sichuan Academy of Natural Resource Sciences, Chengdu 610015, China

15 ⁶U.S. Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for
16 Agriculture and Health, Ithaca, NY 14853, USA

17
18 *W.T., X.S., J.Y. and F.H. contributed equally to this work

19 *Correspondence authors: Dr. Zhangjun Fei, email: zf25@cornell.edu or Dr. Yongsheng Liu,
20 email: liuyongsheng1122@hfut.edu.cn

1
2
3
4 **22 Abstract**

5
6 **23 Background:** Kiwifruit (*Actinidia* spp.) is a dioecious plant with fruits containing abundant
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
vitamin C and minerals. A handful of kiwifruit species have been domesticated, among which the
A. eriantha is increasingly favored in breeding due to its superior commercial traits. Recently, elite
cultivars from *A. eriantha* have been successfully selected and further studies on their biology and
breeding potential require genomic information which is currently unavailable.

28 Findings: Here, we assembled a chromosome-scale genome sequence of *A. eriantha* cv. ‘White’
29 using single-molecular sequencing and chromatin interaction map based scaffolding. The
30 assembly has a total size of 690.6 Mb and an N50 of 21.7 Mb. Approximately 99% of the assembly
31 were in 29 pseudomolecules corresponding to the 29 kiwifruit chromosomes. Forty-three percent
32 of the *A. eriantha* genome are repetitive sequences, and the non-repetitive part encodes 42,988
33 protein-coding genes, of which 39,075 have homologues from other plant species or protein
34 domains. The divergence time between *A. eriantha* and its close relative *A. chinensis* is estimated
35 to be 3.3 million years, and after diversification, 1,727 and 1,506 gene families are expanded or
36 contracted in *A. eriantha*, respectively.

37 Conclusions: We provide a high-quality reference genome for kiwifruit *A. eriantha*. This
38 chromosome-scale genome assembly is substantially better than two published kiwifruit
39 assemblies from *A. chinensis* in terms of genome contiguity and completeness. The availability of
40 *A. eriantha* genome provides a valuable resource for facilitating kiwifruit breeding and the studies
41 of kiwifruit biology.

42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
43 Key words: Kiwifruit; *Actinidia eriantha*; Genome assembly; single molecular sequencing; Hi-C

1
2
3
4 45 **Data description**

5
6 46 ***Introduction***

7
8
9 47 Kiwifruit is well known as the king of fruits due to its remarkably high vitamin C content and
10
11 48 abundant minerals [1, 2]. Native to China, kiwifruit belongs to the genus *Actinidia* which contains
12
13
14 49 54 species and 75 taxa [3]. All species in this genus are perennial, deciduous and dioecious plants
15
16 50 with a climbing or scrambling growth habit, and they also have many common morphological
17
18
19 51 features including the characteristic radiating arrangement of styles of female flower and the
20
21 52 structure of the fruit [4]. Despite rich germplasm resources in kiwifruit, only a few *Actinidia*
22
23
24 53 species have been domesticated, such as *A. chinensis* var. *chinensis*, *A. chinensis* var. *deliciosa* and
25
26 54 *A. eriantha*, whose fruit size are close to commercial standard [5-7].
27

28
29 55 Owing to its strong resistance to *Pseudomonas syringae* pv. *Actinidiae*, long shelf-life,
30
31 56 enriched ascorbic acid and peelable skin [7-11], the *A. eriantha* (2n=58) has been favored in
32
33
34 57 kiwifruit breeding. Recently, new cultivars have been selected either from the wild germplasm of
35
36 58 *A. eriantha* such as ‘White’ (Fig. 1) or from the interspecific hybridization between *A. eriantha*
37
38 59 (♂) and *A. chinensis* (♀) such as ‘Jinyan’ [7, 12]. The ‘White’ has particularly large fruits (96 g
39
40
41 60 on average) with green flesh and favorable flavor and has been widely cultivated in China [7].
42

43 61 *Actinidia eriantha* has also been used for genetic and genomic studies thanks to its high
44
45
46 62 efficiency in genetic transformation and relatively short phase of juvenility [13]. The flowering
47
48
49 63 and fruiting of *A. eriantha* can be accomplished within two years in green house conditions with a
50
51 64 low requirement for winter chilling [13]. In addition, roots of *A. eriantha* which contain many
52
53
54 65 bioactive compounds such as triterpenes and polysaccharides are employed as a traditional
55
56 66 Chinese medicine for the treatment of gastric carcinoma, nasopharyngeal carcinoma, breast
57
58 67 carcinoma, and hepatitis [12, 14].
59
60
61
62
63
64
65

1
2
3
4 68 Previously, two kiwifruit genomes were published and both were from *A. chinensis*
5
6 69 ('Hongyang' and 'Red 5') [15, 16]. These short-read based assemblies are very fragmented,
7
8
9 70 possibly due to the high complexity and heterozygosity of the kiwifruit genomes as well as
10
11 71 technical limitations. Here, we used single-molecular sequencing combined with the high-
12
13
14 72 throughput chromosome conformation capture (Hi-C) technology to assemble the genome of the
15
16 73 elite kiwifruit cultivar 'White' of *A. eriantha*. The availability of this high-quality chromosome-
17
18
19 74 scale genome sequence not only provides fundamental knowledge regarding kiwifruit biology but
20
21 75 also presents a valuable resource for kiwifruit breeding programs.
22

23 24 76 25 26 77 ***Sample collection and genome sequencing***

27
28 78 Fresh young leaves were collected from a female individual of *A. eriantha* cv. 'White'. High
29
30
31 79 molecular weight (HMW) genomic DNA was extracted using the CTAB method as described in
32
33 80 the protocol ([https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Preparing-](https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Preparing-Arabidopsis-DNA-for-20-kb-SMRTbell-Libraries.pdf)
34
35 81 [Arabidopsis-DNA-for-20-kb-SMRTbell-Libraries.pdf](https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Preparing-Arabidopsis-DNA-for-20-kb-SMRTbell-Libraries.pdf)). To construct genomic libraries
36
37
38 82 (SMRTbell libraries) for PacBio long-read sequencing, HMW genomic DNA was sheared into
39
40
41 83 fragments of approximately 20 kb using a Covaris g-Tube (KBiosciences p/n520079),
42
43 84 enzymatically repaired and converted to SMRTbell template following the Manufacturer's
44
45 85 instruction (DNA Template Prep Kit 1.0, PacBio p/n 100-259-100). The templates were size-
46
47
48 86 selected using a BluePippin (SageScience, Inc.) to enrich large DNA fragments (> 10 kb) and then
49
50
51 87 sequenced on a PacBio Sequel system. A total of 9 SMRT cells were sequenced, yielding
52
53 88 3,889,480 million reads with a mean and median length of 10,065 and 15,661 bp, respectively, and
54
55 89 a total of 39.1 Gb sequences, about 52.5× coverage of the kiwifruit genome with an estimated size
56
57
58 90 of 745.3 Mb based on the flow cytometry analysis (Fig. S1; Table S1).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

91 Three paired-end Illumina libraries with insert sizes of 180, 220 and 500 bp, and seven
92 mate-pair libraries with insert sizes of 3, 4, 5, 8, 10, 15, 17 kb, were prepared using Illumina’s
93 Genomic DNA Sample Preparation kit and the Nextera Mate Pair Sample Preparation kit (Illumina,
94 San Diego, CA), respectively. All libraries were sequenced on an Illumina HiSeq 2500 system,
95 which yielded about 80.1 and 97.3 Gb of raw sequence data for paired-end and mate-pair libraries,
96 respectively (Table S1). The raw Illumina paired-end reads were processed to remove duplications,
97 adaptors and low-quality bases using Super-Deduper [17] and Trimmomatic [18] (v0.35), and the
98 mate-pair reads were cleaned using NextClip [19] (v1.3.1) with default parameters. Finally, we
99 obtained 76.6 and 46.2 Gb high-quality cleaned sequences for paired-end and mate-pair libraries,
100 respectively (Table S1).

101 To construct the Hi-C library, ‘White’ plants were grown in a greenhouse, and
102 approximately 4~6 grams young leaves were then harvested and subsequently fixed in the
103 formaldehyde (1% v/v) for 10 min at room temperature. The fixation was terminated by adding
104 glycine to a final concentration of 0.125M. The fixed samples were ground into powder in liquid
105 nitrogen and then lysed with the addition of Triton X-100 to a concentration of 1% (v/v). The
106 nuclei were isolated and prepared for Hi-C library construction according to a previously published
107 protocol [20].

108
109 ***Transcriptome sequencing***

110 To improve gene prediction, we generated transcriptome sequences from a pool of mixed tissues
111 of ‘White’ including root, stem, leaf, flower, and fruits at 7, 30, 60, 90 and 120 days after anthesis.
112 Total RNA was extracted from these tissues using an RNA extraction kit (BIOFIT, China), treated
113 with DNase I and further purified with RNA clean kit (Promega, USA). RNA-Seq libraries were

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

constructed with the NEBNext® Ultra™ RNA Library Prep Kit (Illumina, USA), and sequenced on an Illumina HiSeq 2500 system using paired-end mode. A total of ~19.5 million raw read pairs were obtained, which were processed with Trimmomatic to remove adaptors. The cleaned reads were assembled *de novo* with Trinity [21] (version 2.4.0). Additionally, we also generated genome-guided assemblies with both Trinity and StringTie [22]. Different transcriptome assemblies were eventually integrated by PASA [23] (version 2.3.3) and used as transcript evidence during gene prediction process. Mapping of RNA-Seq reads to the genome assembly was performed with STAR [24] (version 020201), and read counting on the coding regions was performed with HTSeq [25] (version 0.6.0.).

Chromosome-scale assembly of the A. eriantha genome

Actinidia eriantha is a dioecious plant with a heterozygous diploid genome. We estimated the heterozygosity level through the k-mer spectrum analysis with GenomeScope [26] using sequences from the paired-end library with the insert size of 180 bp. The depth distribution of the derived 17-mers clearly showed two separate peaks, based on which we estimated the heterozygosity level of the *A. eriantha* cv. ‘White’ genome to be 1.21% (Fig. S1).

We then estimated the genome size of *A. eriantha* ‘White’ using the flow cytometry analysis, with tomato (*Solanum lycopersicum* cv. Ailsa Craig) used as the reference. We also performed flow cytometry analysis on *A. chinensis* cv. Hongyang. Approximately 1 g of young leaves were washed twice in distilled water and then chopped in ice-cold lysis buffer A (10 mmol/L MgSO₄, 50 mmol/L KCl, 3.5 mmol/L HEPES pH 7.5, 0.3% (v/v) Triton x-100, 2% PVP 30 (W/V)). After 5 minutes, the crude lysate was passed through a 75-µm pore size nylon mesh to remove large cellular debris. The filtrate (1 ml) was transferred to a 1.5 ml plastic tube and

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

centrifuged at 1000 rpm for 5 minutes. The supernatant was discarded, and the nuclei were then resuspended with lysis buffer B (10 mmol/L MgSO₄, 50 mmol/L KCl, 3.5 mmol/L HEPES pH 7.5, 0.3% (v/v) Triton x-100, 0.4 mg/ml Propidium Iodide, 0.04 mg/ml RNase). After 15 minutes, samples were analyzed using a FACS Vantage SE flow cytometer (Becton-Dickinson, San José, USA). Four biological replicates were performed. Based on the 950-Mb genome of tomato, the genome size of ‘White’ was estimated to be 745.3±7.9 Mb, similar to the genome size of *A. chinensis* (Fig. S1) and consistent with that in a previous report (758 Mb; [27]).

We employed a strategy which took into account the unique advantage of different assemblers to construct the ‘White’ genome using PacBio long reads. First, PacBio long reads were corrected and assembled using the Canu program [28] (v1.7), which is a modularized pipeline consisting of three primary stages - read correction, trimming and assembly. The Canu-corrected reads were also assembled independently with the wtdbg program (<https://github.com/ruanjue/wtdbg>), a fast assembler for long noisy reads. Subsequently, the two independent assemblies (one with Canu and another with wtdbg) were merged by Quickmerge [29] (v0.2) to improve the contiguity. The merged assembly was further processed to correct errors using Pilon [30] (version 1.22) with high-quality cleaned Illumina reads from all paired-end and mate-pair libraries representing a total genome coverage of 171× (Table S1). This yielded 2,818,370 nucleotides, 2,495,388 insertions and 1,691,495 deletions being corrected. The resulting final assembled *A. eriantha* cv. ‘White’ genome contained 4,076 contigs with a N50 length of 539,246 bp and a cumulative size of 690,376,929 bp (Table 1). The contiguity and completeness of this assembly far exceeds that of two published kiwifruit *A. chinensis* genomes (Table 1).

160 Table 1 Assembly statistics

	<i>A. eriantha</i>	<i>A. chinensis</i>	
	White	Hongyang	red5
Contigs			
Total contig number (#)	4,076	26,721	39,868
Total contig length (Mb)	690.4	604.2	
Contig N50 (kb)	539.2	58.9	
Contig N90 (kb)	50.7	11.6	
Longest contig length (kb)	3,260.20	423.5	
Scaffolds			
Total scaffold number (#)	1,735	7,698	3,887
Total scaffold length (Mb)	690.6	616.1	550.5
Scaffold N50 (kb)	23,583.9	646.8	623.8
Scaffold N90 (kb)	20,112.1	122.7	140.7
Longest scaffold length (Mb)	28.6	3.4	4.43
Anchored to chromosome (Mb/%)	682.4 / 98.84	452.4 / 73.4	547.9 / 98.9
Anchored with order and orientation (Mb/%)	634.4 / 91.90	333.6 / 54.1	

161

162 To scaffold the contigs based on chromatin interaction maps inferred from the Hi-C data,
163 we first used HiC-Pro [31] to evaluate and filter the cleaned Hi-C reads. The Hi-C data usually
164 contains a considerable part of invalid interaction read pairs which are non-informative and need
165 to be filtered out beforehand. Among the 51 million read pairs that were uniquely aligned to the *A.*
166 *eriantha* assembly, 33 million (64.1%) were valid interaction pairs and their insertion size spanned
167 predominantly from dozens to hundreds of kilobases, therefore providing efficient information for
168 scaffolding. As a part of error correction of the assembly, we used valid Hi-C reads to identify
169 misassembled contigs. In principle, a genuine contig should display a continuous Hi-C interaction
170 map whereas the discrete distribution of an interaction map likely indicates a misassembly. We
171 examined the interaction map for each contig and broke 51 that were possibly misassembled.
172 Subsequently, the corrected PacBio assembly was used for scaffolding with the LACHESIS
173 program [32] and parameters “CLUSTER_MIN_RE_SITES=48,
174 CLUSTER_MAX_LINK_DENSITY=2, CLUSTER_NONINFORMATIVE_RATIO=2,
175 ORDER_MIN_N_RES_IN_TRUN=14, ORDER_MIN_N_RES_IN_SHREDS=15”. LACHESIS

1
2
3
4 176 assigned 3,666 contigs with a total size of 682,355,494 bp (98.84% of the assembly) into 29 groups
5
6
7 177 corresponding to the 29 kiwifruit chromosomes (Fig. 2 and 3a), among which 634,430,648 bp
8
9 178 (91.90%) had defined order and orientation (Table 1 and S2). The final chromosome-scale
10
11
12 179 assembly had a total length of 690,781,529 bp and an N50 of 23,583,865 bp.
13

14 180

15

16 181 ***Evaluation of the genome assembly***

17

18
19 182 We first evaluated the quality of the assembled *A. eriantha* ‘White’ genome by mapping Illumina
20
21 183 genomic and RNA-Seq reads to the assembly. Reads from the paired-end genomic library (with
22
23
24 184 insert size of 180 bp) had very high mapping rate (98.7%), and the properly paired read mapping
25
26 185 rate was 92.0%. For the RNA-Seq reads, 91.7% could be mapped to the genome and 87.1% were
27
28
29 186 uniquely mapped. The high mapping ratio of both genomic and RNA-Seq reads suggest a high
30
31 187 quality of the *A. eriantha* ‘White’ assembly.
32

33
34 188 We then identified synteny between the *A. eriantha* ‘White’ assembly and the assembly of
35
36 189 *A. chinensis* ‘red5’ using MUMMER [33] (version 4.0.0beta2). In general, the two assemblies
37
38 190 showed a high macro-collinearity, with only a few inconsistencies (Fig. 3b). Detailed check of the
39
40
41 191 major inconsistent regions using genetic maps [34] and mate-pair read alignments confirmed the
42
43 192 high quality of the *A. eriantha* ‘White’ genome assembly, and particularly enabled us to discover
44
45 193 that in the ‘red5’ genome a ~8-Mb region was possibly misassembled into chromosome 23 (Fig.
46
47
48 194 S2).
49

50 195

51

52

53 196 ***Repeat annotation***

54

55 197 Repeats were annotated following a protocol described in Campbell et al [35]. The customized
56
57
58 198 repeat library was built to include both known and novel repeat families. We first searched the
59
60
61
62
63
64
65

1
2
3
4 199 assembly for miniature inverted transposable elements (MITEs) using MITE-Hunter [36] with
5
6
7 200 default parameters. The long terminal repeat (LTR) retrotransposons were then identified from the
8
9 201 *A. eriantha* ‘White’ genome using LTRharvest and LTRdigest wrapped in the GenomeTools
10
11 202 package [37]. The LTR identification pipeline was run iteratively to collect both recent (sequence
12
13
14 203 similarity $\geq 99\%$) and old (sequence similarity $\geq 85\%$) LTR retrotransposons. Candidates from each
15
16 204 run were filtered based on the elements typically encoded by LTR retrotransposons. The default
17
18
19 205 parameters (-minlenltr 100 -maxlenltr 6000 -mindistltr 1500 -maxdistltr 25000 -mintsd 5 -maxtsd
20
21 206 5 -motif tgca) were used in LTR calling according to Campbell *et al.* [35]. An initial repeat
22
23
24 207 masking of *A. eriantha* ‘White’ genome was performed with the repeat library derived by
25
26 208 combining the identified MITEs and LTR transposons. The repeat masked genome was fed to
27
28
29 209 RepeatModeler (<http://www.repeatmasker.org/RepeatModeler/>) to identify novel repeat families.
30
31 210 Finally, all identified repeat sequences were combined and searched against a plant protein
32
33
34 211 database where transposon encoding proteins were excluded. Elements with significant similarity
35
36 212 to plant genes were removed. The final repeat library contained 1,670 families, and 526 of them
37
38
39 213 were potentially novel repeat families. We used this species-specific repeat library to mask the *A.*
40
41 214 *eriantha* ‘white’ genome. Approximately 43.3% of the *A. eriantha* ‘White’ genome was masked,
42
43
44 215 and the largest family of repeats was LTR transposons (Table S3). Repeat content identified in *A.*
45
46 216 *eriantha* ‘White’ was much higher than that in *A. chinensis* [e.g. 36% in Hongyang [15]], and this
47
48 217 difference may be largely due to the improvement of the repeat region assembly with PacBio long
49
50
51 218 reads. In addition, divergence between the two kiwifruit species could also contribute to this
52
53 219 difference.

54
55 220

56
57
58 221 ***Prediction and functional annotation of protein-coding genes***

59
60
61
62
63
64
65

1
2
3
4 222 Protein-coding genes were predicted from the repeat-masked *A. eriantha* ‘White’ genome with the
5
6 223 MAKER-P program [35] (version 2.31.10), which integrates evidence from protein homology,
7
8
9 224 transcripts and *ab initio* predictions. The homology-based evidence was derived by aligning
10
11 225 proteomes from 20 plant species to the ‘White’ genome assembly with exonerate (v2.26.1;
12
13
14 226 <https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>). SNAP [38], AUGUSTUS
15
16 227 [39] (version 3.3), and GeneMark-ES [40] (version 4.35) were used for *ab initio* gene predictions.
17
18
19 228 RNA-Seq data generated in this study were assembled *de novo* with Trinity and the assembled
20
21 229 contigs were aligned to the ‘White’ genome assembly to provide transcript evidence. Predictions
22
23
24 230 supported by the three different sources of evidence were finally integrated by MAKER-P, which
25
26 231 resulted in a total of 52,514 primitive gene models. We then filtered and polished these gene
27
28
29 232 models by two steps. First, we combined our RNA-Seq data with others collected from a previous
30
31 233 study [41], and mapped the reads to the ‘White’ genome using the STAR program [24], and a total
32
33 234 of 266 million read pairs were mapped. Based on the mapping, raw count for each predicted gene
34
35
36 235 model was derived and then normalized to CPM (counts per million mapped read pairs). Gene
37
38 236 models with ultra-low expression (CPM < 0.1) were less likely to be real genes. Furthermore, we
39
40
41 237 found that these lowly expressed genes had relatively high annotation edit distance (AED) score,
42
43 238 an indication of low-confidence as defined by MAKER-P program. Therefore, for gene models
44
45
46 239 with CPM < 0.1, we only kept those containing both pfam domains and homologous sequences in
47
48 240 the NCBI nr protein database. After this filtering process 42,751 gene models were kept. Second,
49
50
51 241 the predicted protein-coding genes of kiwifruit *A. chinensis* ‘red5’ have been manually curated
52
53 242 [16], and therefore these gene models should have relatively higher accuracy and could be used to
54
55 243 modify *A. eriantha* ‘White’ gene models whose predictions were not consistently supported by the
56
57
58 244 different types of evidence. To this end, we performed another two *ab initio* predictions using

1
2
3
4 245 BRAKER [42] and GeMoMa [43] (version 1.5.2) with ‘red5’ proteome as the sole evidence. These
5
6 246 two predictions were compared with the gene models predicted by MAKER-P. Consequently, a
7
8
9 247 total of 237 gene models not predicted by MAKER-P were added and another 415 gene models
10
11 248 which had better predictions by BRAKER2 or GeMoMa were used to replace the corresponding
12
13
14 249 gene models predicted by MAKER-P. Finally, we obtained a total of 42,988 protein-coding genes
15
16 250 in the *A. eriantha* ‘White’ genome, with a mean coding sequence (CDS) size of 1,004 bp and
17
18
19 251 containing an average of five exons.

20
21 252 The predicted genes were functionally annotated by blasting their protein sequences against
22
23 253 TAIR [44], Swiss-Prot [45] and TrEMBL [46] databases with an E-value cutoff of 1e-5. Functional
24
25
26 254 descriptions of the protein hits were assembled with the AHRD program
27
28
29 255 (<https://github.com/groupschoof/AHRD>) and transferred to *A. eriantha* genes. Protein domains
30
31 256 were identified using InterProScan [47] (version 5.29-68.0) by searching the protein sequences
32
33 257 against domain databases including PANTHER [48], Pfam [49], SMART [50], and PROSITE [51].
34
35
36 258 The Gene Ontology (GO) terms were assigned to the *A. eriantha* ‘White’ predicted genes using
37
38 259 the Blast2GO program [52] with entries from NCBI protein database and InterProScan.
39
40
41 260 Collectively, 90.9% (N=39,075) of the predicted genes contain at least one annotation from the
42
43 261 above databases (Table S4).

44
45
46 262

47 48 263 ***Evolutionary and comparative genomic analysis***

49
50 264 To infer the divergence time between *A. eriantha* and *A. chinensis*, we identified gene orthology
51
52
53 265 between the two species using MCScanX [53] and calculated synonymous substitution rate (Ks)
54
55 266 between each orthologous pair. Three additional species, cultivated tomato (*Solanum*
56
57
58 267 *lycopersicum*), wild tomato (*S. pennellii*) and potato (*S. tuberosum*), were also included in the
59
60
61
62
63
64
65

1
2
3
4 268 analysis. The Ks distribution (Fig. 4a) suggested that the divergence between the two kiwifruit
5
6 269 species was earlier than that between the two tomato species. We dated the divergence by assuming
7
8
9 270 a strict molecular clock [54], and the time when *A. eriantha* and *A. chinensis* separated was
10
11 271 estimated to be ~3.3 million years ago (Mya), compared to ~1.9 Mya between *S. lycopersicum* and
12
13
14 272 *S. penellii* and ~6.0 Mya between *S. lycopersicum* and *S. tuberosum*. Gene family evolution was
15
16 273 analyzed by comparing genomes of *A. eriantha*, *A. chinensis*, *S. lycopersicum*, *S. tuberosum*, *Vitis*
17
18
19 274 *vinifera*, *Arabidopsis thaliana* and *Oryza sativa*. A total of 17,593 orthogroups were defined by
20
21 275 OrthoFinder [55] (version 2.2.6) and among which 1,246 were single-copy gene families (Fig. 4b).
22
23
24 276 The single-copy family genes were aligned and concatenated to build a species phylogenetic tree
25
26 277 using IQ-TREE [56] (version 1.5.5) with a best-fitting model (Fig. 4c). Gene family
27
28
29 278 expansion/contraction along the branches of the phylogenetic tree was analyzed by CAFÉ [57]
30
31 279 (version 4.1). Finally, a total of 1,727 and 1,506 gene families were found apparently expanded
32
33 280 and contracted, respectively, in *A. eriantha* (Fig. 4c).
34
35
36 281

37 38 282 **Conclusion**

39
40 283 Here, we report a high-quality reference genome of kiwifruit *A. eriantha* cv. ‘White’. The assembly
41
42
43 284 from single-molecular sequencing combined with Hi-C scaffolding yielded a highly continuous
44
45
46 285 and complete genome than the two previously published kiwifruit genomes. This genome will
47
48 286 provide a valuable source for exploration of genetic basis of unique traits in kiwifruit and also
49
50 287 facilitate the studying of sexual determination loci in the dioecious plants.
51
52

53 288 54 55 289 **Availability of supporting data**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

290 This Whole Genome Shotgun project has been deposited at DBJ/ENA/GenBank under the
291 accession QOVS000000000. The version described in this paper is version QOVS01000000. Raw
292 sequencing reads have been deposited in the Sequence Read Archive (SRA) database under the
293 accession number SRP155011. The *Actinidia eriantha* ‘White’ genome sequence and the
294 annotation are also available at Kiwifruit Information Resource (<http://bdg.hfut.edu.cn/kir/>).
295 Detailed protocols of computational analyses have been deposited in protocols.io:
296 <http://dx.doi.org/10.17504/protocols.io.vgse3we>.

298 Competing interests

299 The authors have no competing interests to declare.

301 Abbreviation

- 302 Blast: Basic Local Alignment Search Tool;
- 303 CTAB: Cetyl trimethylammonium bromide;
- 304 NCBI: National Center for Biotechnology Information;
- 305 RNA-Seq: RNA sequencing;
- 306 PacBio: Pacific Biosciences;
- 307 SMRT: Single Molecule Real-Time;
- 308 Mb: megabase;
- 309 Gb: gigabase

311 Acknowledgement

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

312 This work was supported by grants from the National Natural Science Foundation of China
313 (31471157 and 31700266), National Foundation for Germplasm Repository of Special
314 Horticultural Crops in Central Mountain Areas of China (NJF2017-69), National Science Fund for
315 Distinguished Young Scholars (30825030), Key Project from the Government of Sichuan Province
316 (2013NZ0014, 2016NZ0105), Key Project from the Government of Anhui Province
317 (2012AKKG0739;1808085MC57), and the US National Science Foundation (IOS-1339287 and
318 IOS-1539831).

319

Author contribution

320
321 W.T., X.S. and J.Y. contributed equally to this work. W.T., J.Y., X.T., Y.Y., X.N., M.M., D.Z.,
322 S.H., W.S., C.F. and M.L. collected plant samples, extracted DNA/RNA, and performed
323 transcriptome sequencing and gene expression analyses; W.T., X.S., J.Y., X.T., C.J., Z.F. and Y.L.
324 performed DNA sequencing, genome assembly, gene annotation, evolution and comparative
325 genomic analyses, and website construction; X.S., W.T., Z.F. and Y.L. wrote and revised the
326 manuscript; Y.L. and Z.F. conceived strategies, designed experiments and managed projects. All
327 authors read and approved the manuscript.

328

329

Figure legends

330
331 **Figure 1.** Tree and fruits of *A. eriantha* cv. ‘White’.

332

333 **Figure 2.** Chromatin interaction map of *A. eriantha* derived from Hi-C data. Each group represents
334 an individual chromosome.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

335

Figure 3. Genome of *A. eriantha* and synteny between the two kiwifruit species. (a) Genome landscape of *A. eriantha* cv. ‘White’. Track A: gene density, Track B: repeat density, Track C: GC content; all were calculated in a 500-kb window; (b) Genome synteny between *A. eriantha* cv. ‘White’ and *A. chinensis* ‘red5’.

340

Figure 4. Evolutionary and comparative genomic analyses. (a) Distribution of synonymous substitution rate (Ks) between *A. eriantha* and *A. chinensis*, *S. lycopersicum* and *S. penellii*, and *S. lycopersicum* and *S. tuberosum*; (b) Orthogroups shared by selected species; (c) Species phylogenetic tree and gene family evolution. Numbers on the branch indicate counts of gene family that under either expansion (red) or contraction (green).

346

Figure S1. Genome characteristics of *A. eriantha* and *A. chinensis*. (a) Flow cytometry analyses of *A. eriantha* cv. White and *A. chinensis* cv. Hongyang. The main peak (I) indicates G0/G1 cells and the secondary peak (II) represents G2/M cells. (b) Flow cytometry analyses of *A. eriantha* ‘White’ and *Solanum lycopersicum* cv. Ailsa Craig. Peaks a and b represent the G0/G1 cells of “White” and ‘Ailsa Craig’, respectively. The genome size of ‘White’ was estimated to be 745.3±7.9 Mb using ‘Ailsa Craig’ as the reference. (c) 17-mer distribution of ‘White’ genomic reads (180bp paired-end library).

354

Figure S2. Examination of assembly inconsistencies between *A. eriantha* cv. ‘White’ and *A. chinensis* cv. ‘red5’. (a) Validation of genome assembly of ‘White’ using genetic maps. Horizontal lines within “White” chromosomes indicate gapped regions and lines between chromosomes of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

358 two assemblies indicate syntenic regions. (b) A chromosomal segment assembled into the Chr23
359 in *A. chinensis* “red5”, is syntenic to the region located at the terminus of Chr19 in *A. eriantha*
360 cv. ‘White’. (c) Snapshots of Illumina mate-pair reads mapped to the junctions of the break point
361 as well as nearby regions supporting the assembly of ‘White’.

362

Reference

- 1 Ferguson AR, Ferguson LR. Are kiwifruit really good for you? *Acta Hort* 2013;**610**:131-138
- 2 Richardson DP, Ansell J, Drummond LN. The nutritional and health attributes of kiwifruit: a review. *Eur J Nutr* 2018;1-18.
- 3 Li JQ, Li XW, Soejarto DD. Actinidiaceae. In: Wu ZY, Raven PH, Hong DY, eds. *Flora of China*. Beijing: Science Press & St. Louis: Missouri Plant Garden Press; 2007;**12**:pp334-362.
- 4 Ferguson AR, Huang H. Genetic resources of kiwifruit: domestication and breeding. *Hortic Rev.* 2007;**33**:1-121.
- 5 Testolin R. Kiwifruit (*Actinidia* spp.) in Italy: The history of the industry, international scientific cooperation and recent advances in genetics and breeding. *ISHS Acta Horticulturae* 2015;47-61.
- 6 Jo YS, Cho HS, Park MY, Bang GP. Selection of a sweet *Actinidia eriantha* 'bidan'. *ISHS Acta Horticulturae* 2017; 253-258.
- 7 Wu Y, Xie M, Zhang Q et al. Characteristics of 'White': a new easy-peel cultivar of *Actinidia eriantha*. *N Z J Crop Hortic Sci* 2009;**37**(4):369-373.
- 8 Atkinson RG, Sharma NN, Hallett IC et al. *Actinidia eriantha*: a parental species for breeding kiwifruit with novel peelability and health attributes. *N Z J For Sci* 2009;**39**:207-216.
- 9 Guo R, Landis JB, Moore MJ et al. Development and application of transcriptome-derived microsatellites in *Actinidia eriantha* (Actinidiaceae). *Front Plant Sci* 2017;**8**:1383.
- 10 Prakash R, Hallett IC, Wong SF et al. Cell separation in kiwifruit without development of a specialised detachment zone. *BMC Plant Biol.* 2017;**17**(1):86.
- 11 Shi ZJ, Zhang HQ, Hui Q et al. The resistance evaluation of different kiwifruit varieties to canker. *Acta Agriculturae Zhejiangensis* 2014;**26**(3):752-759
- 12 Zhang D, Gao C, Li R et al. TEOA, a triterpenoid from *Actinidia eriantha*, induces autophagy in SW620 cells via endoplasmic reticulum stress and ROS-dependent mitophagy. *Arch Pharm Res* 2017;**40**(5):579-591.
- 13 Wang T, Ran Y, Atkinson RG et al. Transformation of *Actinidia eriantha*: a potential species for functional genomics studies in Actinidia. *Plant Cell Rep.* 2006;**25**(5):425-431.
- 14 Wu JG, Ma L, Lin SH et al. Anticancer and anti-angiogenic activities of extract from *Actinidia eriantha* Benth root. *J Ethnopharmacol* 2017;**203**:1-10.
- 15 Huang S, Ding J, Deng D et al. Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun* 2013;**4**:2640.
- 16 Pilkington SM, Crowhurst R, Hilario E 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**(1):257.
- 17 Petersen KR, Streett DA, Gerritsen AT et al. Super deduper, fast PCR duplicate detection in fastq files. *ACM* 2015;491-492.
- 18 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114-20.
- 19 Leggett RM, Clavijo BJ, Clissold L et al. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* 2013;**30**(4):566-568.
- 20 Rao SS, Huntley MH, Durand NC et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;**159**(7):1665-1680.
- 21 Haas BJ, Papanicolaou A, Yassour M et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013;**8**(8):1494.
- 22 Pertea M, Pertea GM, Antonescu CM et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015;**33**(3):290-295.
- 23 Haas BJ, Delcher AL, Mount SM et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 2003;**31**(19):5654-5666.
- 24 Dobin A, Davis CA, Schlesinger F et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**(1):15-21.

1
2
3
4 413 25. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput
5 414 sequencing data. *Bioinformatics* 2015;**31**(2):166-169.

6 415 26. Vurture GW, Sedlazeck FJ, Nattestad M et al. GenomeScope: fast reference-free genome profiling
7 416 from short reads. *Bioinformatics* 2017;**33**(14):2202-2204.

8 417 27. Hopping ME. Flow cytometric analysis of *Actinidia* species. *N Z J Bot* 1994;**32**(1):85-93.

9 418 28. Koren S, Walenz BP, Berlin K et al. Canu: scalable and accurate long-read assembly via adaptive
10 419 k-mer weighting and repeat separation. *Genome Res* 2017;**27**(5):722-736.

11 420 29. Chakraborty M, Baldwin-Brown JG, Long AD et al. Contiguous and accurate de novo assembly of
12 421 metazoan genomes with modest long read coverage. *Nucleic Acids Res* 2016;**44**(19):e147.

13 422 30. Walker BJ, Abeel T, Shea T et al. Pilon: an integrated tool for comprehensive microbial variant
14 423 detection and genome assembly improvement. *PloS One* 2014;**9**(11):e112963.

15 424 31. Servant N, Varoquaux N, Lajoie BR et al. HiC-Pro: an optimized and flexible pipeline for Hi-C
16 425 data processing. *Genome Biol* 2015;**16**(1):259.

17 426 32. Burton JN, Adey A, Patwardhan RP et al. Chromosome-scale scaffolding of de novo genome
18 427 assemblies based on chromatin interactions. *Nat Biotechnol* 2013;**31**(12):1119.

19 428 33. Kurtz S, Phillippy A, Delcher AL et al. Versatile and open software for comparing large genomes.
20 429 *Genome Biol* 2004;**5**(2):R12.

21 430 34. Zhang Q, Liu C, Liu Y et al. 2015. High-density interspecific genetic maps of kiwifruit and the
22 431 identification of sex-specific markers. *DNA Res* 2015;**22**(5):367-375.

23 432 35. Campbell M, Law M, Holt C et al. MAKER-P: a tool-kit for the rapid creation, management, and
24 433 quality control of plant genome annotations. *Plant Physiol* 2013;**164**(2):513-524.

25 434 36. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat
26 435 transposable elements from genomic sequences. *Nucleic Acids Res* 2010;**38**(22):e199.

27 436 37. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient
28 437 processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform*
29 438 2013;**10**(3):645-656.

30 439 38. Korf I. Gene finding in novel genomes. *BMC Bioinformatics* 2004;**5**(1):59.

31 440 39. Stanke M, Keller O, Gunduz I et al. AUGUSTUS: ab initio prediction of alternative transcripts.
32 441 *Nucleic Acids Res.* 2006;**34**:W435-W439.

33 442 40. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO et al. Gene identification in novel eukaryotic
34 443 genomes by self-training algorithm. *Nucleic Acids Res* 2005;**33**(20):6494-6506.

35 444 41. Wang Z, Liu Y, Li D et al. Identification of circular RNAs in kiwifruit and their species-specific
36 445 response to bacterial canker pathogen invasion. *Front Plant Sci.* 2017;**8**:413.

37 446 42. Hoff KJ, Lange S, Lomsadze A et al. BRAKER1: Unsupervised RNA-Seq-Based Genome
38 447 Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 2016;**32**(5):767-769.

39 448 43. Keilwagen J, Wenk M, Erickson JL et al. Using intron position conservation for homology-based
40 449 gene prediction. *Nucleic Acids Res* 2016;**44**(9):e89.

41 450 44. Rhee SY, Beavis W, Berardini TZ et al. The Arabidopsis Information Resource (TAIR): a model
42 451 organism database providing a centralized, curated gateway to Arabidopsis biology, research
43 452 materials and community. *Nucleic Acids Res* 2003;**31**(1):224-228.

44 453 45. Bairoch A, Boeckmann B. The SWISS-PROT protein sequence data bank. *Nucleic Acids Res*
45 454 1991;**19**(Suppl):2247-2249.

46 455 46. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement
47 456 TrEMBL. *Nucleic Acids Res* 1997;**25**(1):31-36.

48 457 47. Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition
49 458 methods in InterPro. *Bioinformatics* 2001;**17**(9):847-848.

50 459 48. Mi H, Lazareva-Ulitsky B, Loo R et al. The PANTHER database of protein families, subfamilies,
51 460 functions and pathways. *Nucleic Acids Res* 2005;**33**(Suppl):D284-D288.

52 461 49. Finn RD, Bateman A, Clements J et al. Pfam: the protein families database. *Nucleic Acids Res*
53 462 2014;**42**(Database issue):D222-D230.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

463 50. Schultz J, Copley RR, Doerks T et al. SMART: a web-based tool for the study of genetically mobile
464 domains. *Nucleic Acids Res* 2000;**28**(1):231-234.
465 51. Bairoch A. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res*
466 1991;**19**(Suppl):2241-2245.
467 52. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int*
468 *J Plant Genomics* 2008;**2008**:619832.
469 53. Wang Y, Tang H, DeBarry JD et al. MCScanX: a toolkit for detection and evolutionary analysis of
470 gene synteny and collinearity. *Nucleic Acids Res* 2012;**40**(7):e49.
471 54. Ossowski S, Schneeberger K, Lucas-Lledó JI et al. The rate and molecular spectrum of spontaneous
472 mutations in *Arabidopsis thaliana*. *Science* 2010;**327**(5961):92-94.
473 55. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons
474 dramatically improves orthogroup inference accuracy. *Genome Biol* 2015;**16**(1):157.
475 56. Nguyen LT, Schmidt HA, von Haeseler A et al. IQ-TREE: a fast and effective stochastic algorithm
476 for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2014;**32**(1):268-74.
477 57. De Bie T, Cristianini N, Demuth JP et al. CAFE: a computational tool for the study of gene family
478 evolution. *Bioinformatics* 2006;**22**(10):1269-71.
479



Figure 1

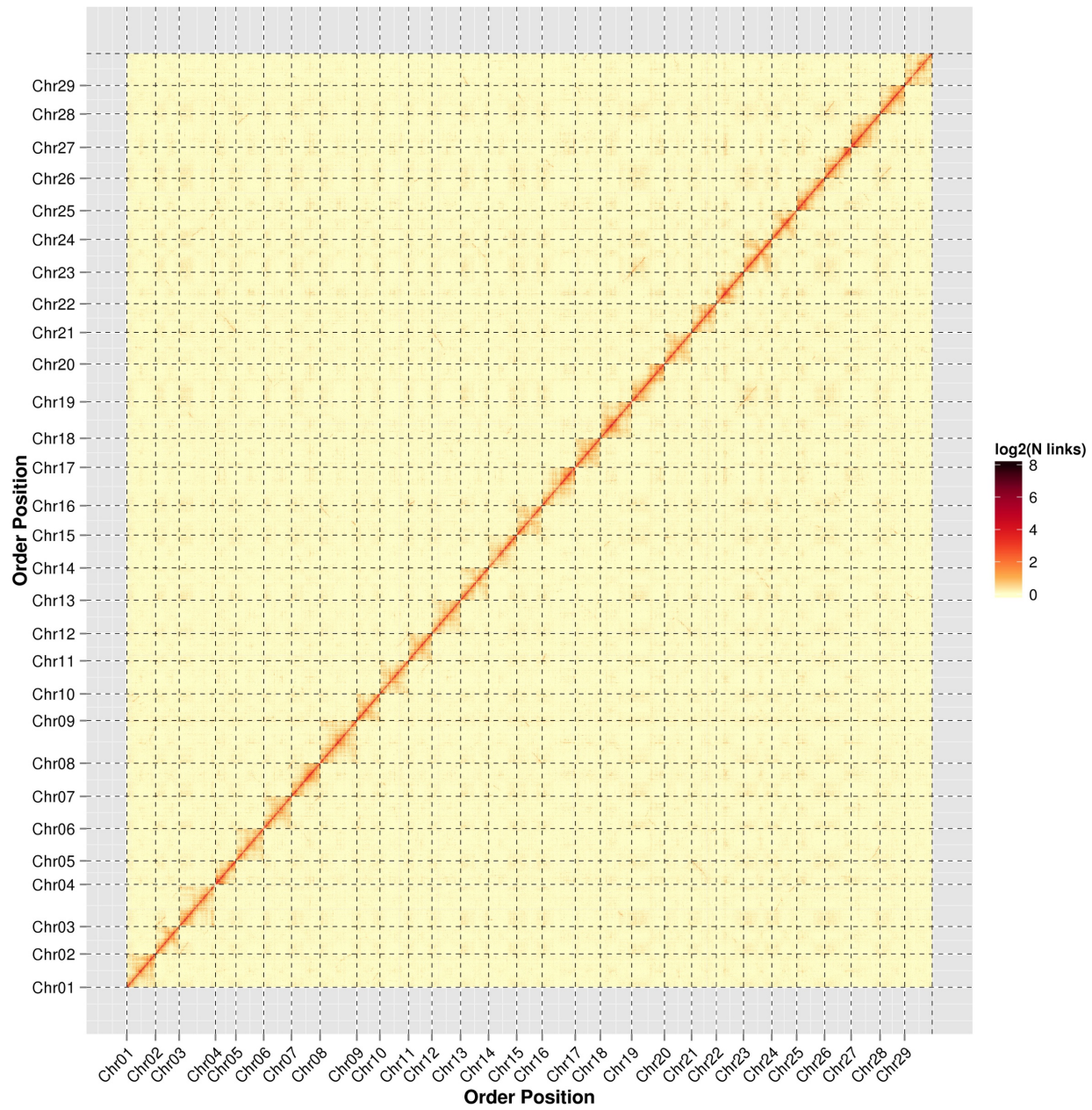
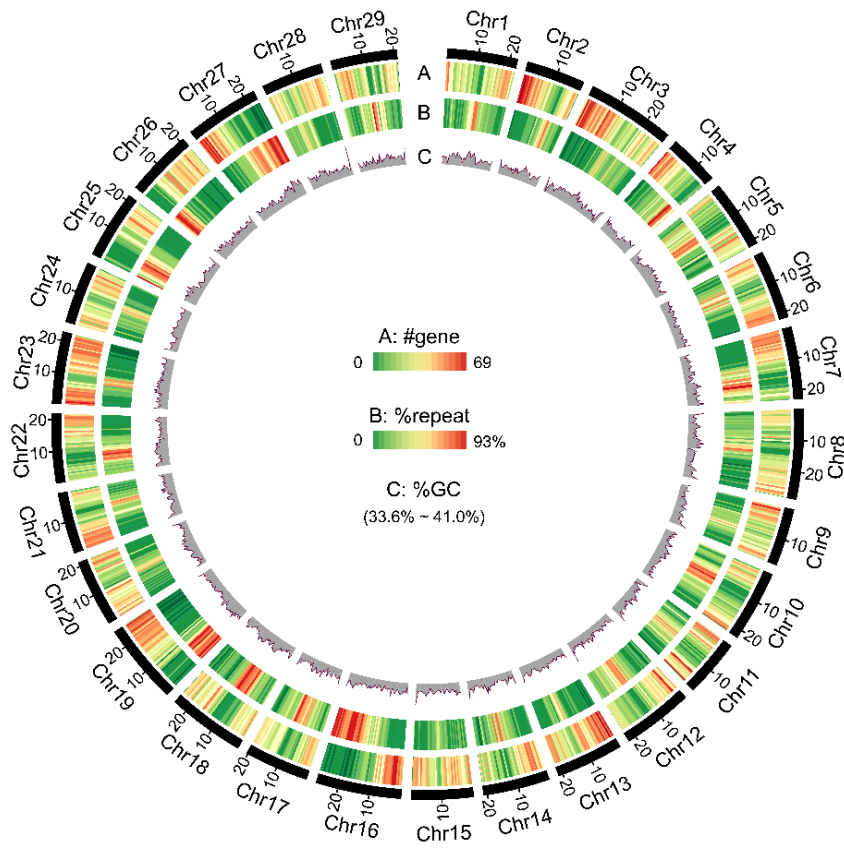


Figure 2.

a



b

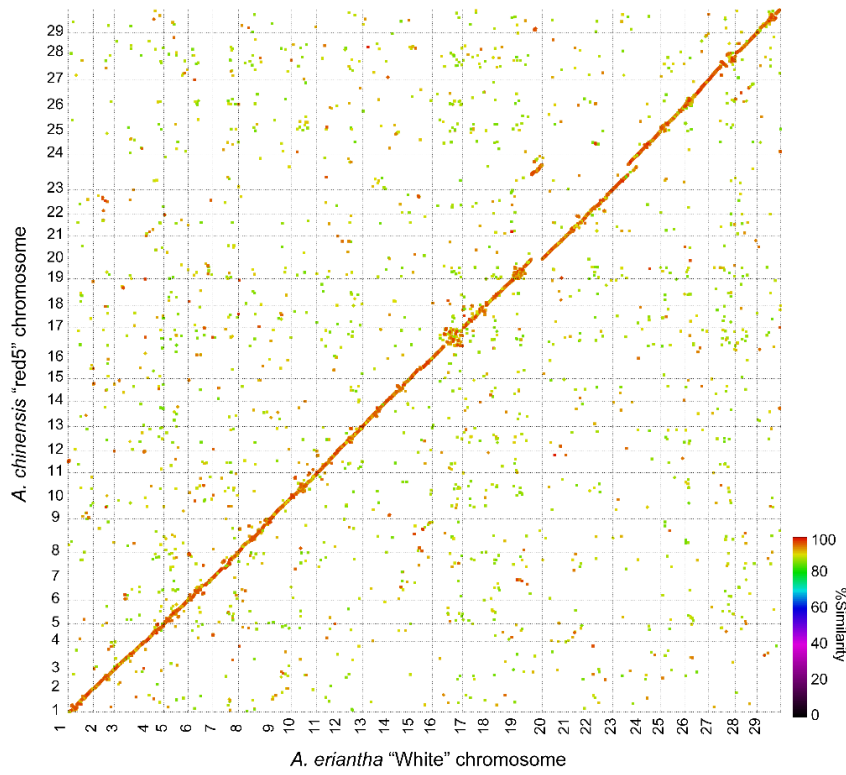


Figure 3

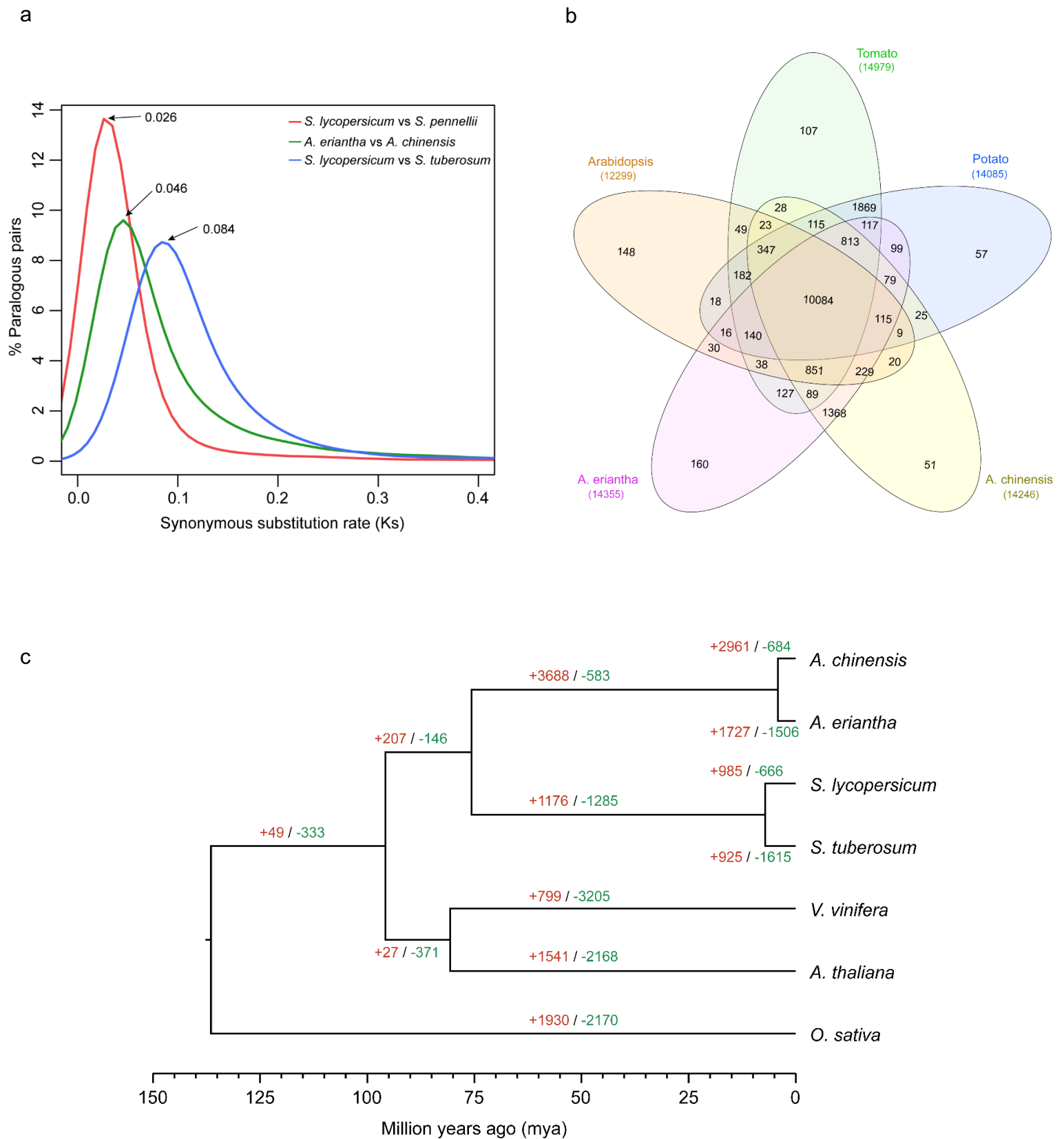
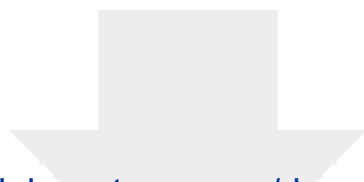


Figure 4

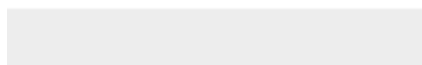
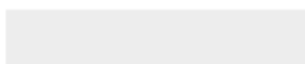


Click here to access/download
Supplementary Material
Supp_Figures.pdf



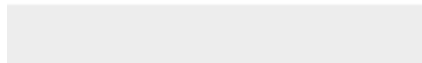


Click here to access/download
Supplementary Material
Supp_Tables.xlsx





Click here to access/download
Supplementary Material
Kiwifruit_response_letter.pdf



December 31, 2018

Dear Dr. Hongling Zhou

Thank you again for your time and effort in handling our manuscript titled “Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin conformation capture” (manuscript reference number: GIGA-D-18-00282R1).

We have attached our detailed responses to the specific comments of the reviewer 1. We hope that we have addressed the reviewer 1’s concerns and ask that you please consider this revised manuscript for publication in GigaScience.

Sincerely,

Zhangjun Fei