nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0506-8

In the format provided by the authors and unedited.

The *bracteatus* pineapple genome and domestication of clonally propagated crops

Li-Yu Chen[1,21], Robert VanBuren[2,3,21], Margot Paris[4,21], Hongye Zhou[5,21], Xingtan Zhang[1], Ching Man Wai[2], Hansong Yan[1], Shuai Chen[1], Michael Alonge[6], Srividya Ramakrishnan[6], Zhenyang Liao[1], Juan Liu[1], Jishan Lin[1], Jingjing Yue[1], Mahpara Fatima[1], Zhicong Lin[1], Jisen Zhang[1], Lixian Huang[1], Hao Wang[5], Teh-Yang Hwa[7], Shu-Min Kao[7], Jae Young Choi[8], Anupma Sharma[9], Jian Song[10], Lulu Wang[1], Won C. Yim[11], John C. Cushman[11], Robert E. Paull[12], Tracie Matsumoto[13], Yuan Qin[1], Qingsong Wu[14], Jianping Wang[1,10], Qingyi Yu[1,9], Jun Wu[15], Shaoling Zhang[15], Peter Boches[13], Chih-Wei Tung[7], Ming-Li Wang[16], Geo Coppens d'Eeckenbrugge[17,18], Garth M. Sanewski[19], Michael D. Purugganan[8], Michael C. Schatz[6], Jeffrey L. Bennetzen[5]*, Christian Lexer[20]* and Ray Ming[1,2]*

---

[1]FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou, China. [2]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Department of Horticulture, Michigan State University, East Lansing, MI, USA. [4]Department of Biology, University of Fribourg, Fribourg, Switzerland. [5]Department of Genetics, University of Georgia, Athens, GA, USA. [6]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [7]Department of Agronomy, National Taiwan University, Taipei, ROC. [8]Department of Biology, Center for Genomics and Systems Biology, New York University, NY, New York, USA. [9]Texas A&M AgriLife Research, Texas A&M University System, Dallas, TX, USA. [10]Department of Agronomy, University of Florida, Gainesville, FL, USA. [11]Department of Biochemistry and Molecular Biology, MS330, University of Nevada, Reno, NV, USA. [12]Department of Tropical Plant and Soil Sciences, University of Hawaii at Manoa, Honolulu, HI, USA. [13]USDA-ARS, Pacific Basin Agricultural Research Center, Hilo, HI, USA. [14]South Subtropical Crops Research Institute, CATAS, Zhanjiang, China. [15]Centre of Pear Engineering Technology Research, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China. [16]Hawaii Agriculture Research Center, Kunia, HI, USA. [17]Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR AGAP, Montpellier, France. [18]AGAP, Université de Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. [19]Queensland Department of Agriculture and Fisheries, Nambour, Queensland, Australia. [20]Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria. [21]These authors contributed equally: Li-Yu Chen, Robert VanBuren, Margot Paris and Hongye Zhou. *e-mail: maize@uga.edu; christian.lexer@univie.ac.at; rayming@illinois.edu

The *bracteatus* pineapple genome and domestication of clonally propagated crops

Supplementary Note

Plant materials, DNA/RNA extraction, and library construction. The CB5 plants used for DNA extraction were grown and collected in green house at Fujian Agriculture and Forestry University, Fuzhou, China. The CB5 leaf DNA for PacBio sequencing was isolated following the procedure described previously[1]. Basal white section of leaves from the 78 domesticated pineapples, 9 wild *Ananas* accessions, and two *Pitcairnia* accessions as outgroup were collected for DNA isolation, and among them, 76 were provided by the USDA tropical plant germplasm collection in Hilo, Hawaii and 13 were provided by the Ministry of agriculture Pineapple Germplasm Repository of Zhanjiang city at South Subtropical Crops Research Institute, CATAS China. White leaf tissue was chosen to reduce plastid contamination. Genomic DNA for whole genome resequencing samples was extracted using the DNAeasy plant mini kit (Qiagen). Paired-end DNA-seq libraries with an average insert size of 500 bp were made using the Illumina TruSeq DNA LT kit (ID: FC-121-2001) according to the manufacturer's instructions (Illumina). DNAseq libraries were sequenced on an Illumina HiSeq2500 system under paired-end mode with 100-bp, 150-bp, or 250-bp reads.

The CB5 plants used for RNA extraction were grown and collected in Kunia field station on Oahu, Hawaii at the Hawaii Agriculture Research Center. For CB5 leaf and flower samples used for RNA extraction, mature leaf and mixed stages of flowers were collected from three CB5 plants and three biological replicates were collected. For CB5 fruits used for RNA extraction, we collected 5 fruits that were observed one month after flower primordial initiation, and 5 fruits observed two months after flower primordial initiation. Each fruit serves as a biological replicate. The ‘F153’ plants used for flower, root, and mature leaf RNA extraction were grown and collected in Kunia field station on Oahu, Hawaii at the Hawaii Agriculture Research Center. The ‘F153’ flower, root, and mature leaf were collected from 3 plants and only 1 biological replicate was sequenced. For ‘F153’ mature androecium and gynoecium, these two tissues were collected from pineapple plants in green house at Fujian Agriculture and Forestry University, Fuzhou, China. Plants were grown in green house with 12 h light at 28 ℃ followed by 12 h dark at 22 ℃. All tissues were hand-dissected and frozen immediately in liquid nitrogen. The tissues from at least three independent inflorescences were combined into one biological replicate and there were three biological replicates for each sample. For MD2 fruit, 8 stages of fruits were collected from the field of Dole Plantation, Wahiawa, Hawaii for RNA extraction for the fruit ripening developmental series. Total RNA was extracted from aforementioned tissues using the Qiagen RNeasy Plant Mini Kit (Qiagen, #74904) following the manufacturer’s protocol. DNA contamination was removed using the DNA-*free*™ DNA Removal Kit (Life Technologies, #AM1906M). RNA-seq libraries were constructed using the Illumina TruSeq stranded RNA Sample Preparation Kit

(Illumina, #RS-122-2001) and the libraries were sequenced on an Illumina HiSeq2500 System under paired- or single-end mode. Detailed information on all RNA sequencing samples are listed in Supplementary Table 26.

Assembly and Removal of Heterozygous Contigs in CB5 Genome. The raw PacBio reads were error corrected and assembled using Canu v1.7[2] with the following optimized parameters, which allowed for more careful unitigging (batOptions):

```
canu-1.7/Linux-amd64/bin/canu \
   -p cb5 -d canu genomeSize=650m \
   gridEngineThreadsOption="-pe threads THREADS" \
   gridEngineMemoryOption="-l m_mem_free=MEMORY" \
   corOutCoverage=200 "batOptions=-dg 3 -db 3 -dr 1 -ca 500 -cp 50" \
   minReadLength=12000 \
   -pacbio-raw *.fa.gz
```

The total span of this assembly was 809.6 Mb, with a contig N50 size of 280.7 kb. Bacterial contamination were identified and filtered out by mapping the contigs to the COGE v6 reference pineapple assembly[3], as well as all bacterial contigs in NCBI Refseq[4] using minimap2[5]. Chimeric contigs that contained both pineapple genomes and baterial sequences were removed if more of their bases aligned to the bacterial genomes than the pineapple genome. A total of ~39 Mb (4.8% of bases) of bacterial contamination were removed from the assembly, leaving a total span of 770.1 Mb and a contig N50 size of 283 kb. Importantly, this cleaning procedure did not decrease the number of eukaryotic core genes present as reported by BUSCO[6].

After cleaning bacterial contamination, the total span of the assembly was still nearly twice the expected (haploid) genome size, which is indicative of the assembler partially resolving the heterozygosity. This hypothesis was further confirmed by a high rate of BUSCO duplicated genes: 48.7% of the BUSCO genes were duplicated in the assembly, compared to only 5.8% in the COGE reference. Consequently, we further processed the assembly to create a pseudo-haploid representation where pairs of contigs representing homologous sequences were filtered to select only one representative contig. Our approach is similar to the approach used by FALCON-unzip[7] for PacBio reads or SuperNova[8] for 10 × Genomics Linked Reads. As with those algorithms, our algorithm will not necessarily maintain the same phase throughout the assembly, and can arbitrarily alternate between homologous chromosomes at the ends of contigs (Supplementary Fig. 22). Unlike those methods, our method can be run stand-alone with any assembler and is available open-source in github at http://github.com/schatzlab/pseudohaploid.

Briefly, the algorithm begins by aligning the genome assembly to itself using the whole genome aligner nucmer from the MUMmer suite[9]. We used the parameters “nucmer -maxmatch -l 100 -c 500” to report all alignments, unique and repetitive, at least 500 bp long with a 100 bp seed match. We further filtered these alignments to those that are 1000 bp or longer using delta-filter (also part of the MUMmer suite). We used and recommend the sge_mummer version of MUMmer so the alignments can be computed in parallel in a cluster environment (https://github.com/fritzsedlazeck/sge_mummer) although this will produce identical results to the serial version. Finally, we filtered the alignments to keep those that were at 90% identity or greater. This filters lower identity repetitive alignments while accommodating the expected rate of heterozygosity between homologous chromosomes while also accounting for local regions of greater diversity. We achieved similar results using slightly higher or lower rates of alignment identity.

Next, the alignments were examined to identify and filter out redundant homologous contigs. As the alignments will contain some repetitive alignments plus the homologous alignments of interest, we removed the spurious repetitive alignments by identifying and focusing on long “alignment chains” consisting of sets of alignments that are co-linear along the pair of contigs (Supplementary Fig. 23). Our method was inspired by older methods for computing synteny between distantly related genomes using alignment chains[10], although our method is more focused for this problem. As we expect there to be structural variations between the homologous sequences, we allow for gaps in the alignments between the contigs, although true homologous contig pairs should maintain a consistent order and orientation to the alignments. Specifically, in the alignments from contig A to contig B, each aligned region of A forms a node in an alignment graph, and edges are added between nodes if they are compatible alignments, meaning they are on the same strand, and the implied gap distance on both contig A and contig B was less than 20 kb but not negative. Our algorithm then uses a depth first search starting at every node in the alignment graph to find the highest scoring chain of alignments, where the score is determined by the number of bases that are aligned in the chain. Notably, if a repetitive alignment is flanked by unique or repetitive alignments, such as the orange sequence in Contig B below, this approach will prefer to chain together nearby alignments that are co-linear on Contig A. We find this produces more accurately identify homologous contigs than the filtering that MUMmer’s delta-filter can perform, which does not consider the context of the alignments and therefore often selects suboptimal repetitive alignments[11].

With the alignment chains identified between pairs of contigs, the last phase of the algorithm is to remove any contigs that are redundant with other contigs originating on the homologous chromosome. Specifically, it evaluates the contigs in order from smallest to longest, and computes the fraction of the bases of each contig that are spanned by alignment chains to other non-redundant contigs. If more than 93% of the contig bases are spanned, it is marked as redundant. This can occur in simple cases where shorter contigs are spanned by individual longer contigs **(**Supplementary Fig. 24a**)** as well as more complex cases where a contig is spanned by multiple shorter non-redundant contigs (Supplementary Fig. 24b). We evaluated several cutoffs for the threshold of percent of the bases spanned, and ultimately selected 93% as this minimized the number of duplicated BUSCO genes while not substantially increasing the number of missing BUSCO genes. The total span of the final filtered assembly was 513,048,691 bp, in 1,970 contigs with a contig N50 size of 426,696 bp.

Validation of Heterozygous Contig Assembly Methods. To further demonstrate the parameters used for Canu and the capabilities of our new Pseudohaploid method, we applied these techniques to a highly heterozygous sample of *Arabidopsis thaliana*, an F1 hybrid of Col-0 and Cvi-0 that was previously sequenced as part of the FALCON-unzip paper[7]. For this analysis, we downloaded 116 × coverage of PacBio reads (read N50 length = 17,474) of the F1 genome from the SRA under accession SRX1715706. We then assembled the reads using Canu using the same parameters as described above for the CB5 assembly. As with CB5, the total size of the raw Canu assembly was substantially larger than the expected haploid genome size: the total assembly size was 214.7 Mb, whereas the haploid genome size is ~135 Mb according to the latest estimates from The Arabidopsis Information Resource (TAIR) (https://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/agicomplete.jsp).

We then applied the Pseudohaploid method using the same parameters as we used for the CB5 assembly. Similar to what we observed in CB5, this reduced the total size of the assembly from 214.7 Mb to 143.5 Mb, and increased the contig N50 size from 350 kb to 950 kb by reducing the number of contigs from 2074 to 505. Then using the high quality TAIR10 reference genome, we investigated the quality of both the raw and Pseudohaploid assemblies. Using BUSCO, we found the reference genome contained 1356 complete BUSCOs genes, of which 1348 were single-copy, and 8 were duplicated. Similar to CB5, we found the raw Canu assembly contained a large fraction of duplicated genes, and overall it contained 1355 complete BUSCOs, although only 711 were single-copy, and 644 were duplicated. In contrast the Pseudohaploid assembly substantially reduced the number of duplicate genes, and contained a total of 1355 complete BUSCOs, of which 1240 were single-copy, and only 115 duplicated (an 83% reduction).

Furthermore, by aligning the raw Canu and Pseudohaploid assemblies to the reference TAIR10 assemblies using nucmer using the parameters “-maxmatch -l 100 -c 500”, we found that 1.6 Mb (1.4%) of the TAIR10 assembly was not represented in the Canu assembly, and 4.2 Mb (3.5%) was not represented in the Pseudohaploid assembly as computed by the MUMmer tool dnadiff in the “AlignedBases” field. We also found that 19.0 Mb of the raw Canu assembly and 14.1 Mb of the Pseudohaploid assembly were unaligned to the reference genome.

However, the reference TAIR10 assembly was assembled from the Col-0 accession, and the portions that do not align are chiefly due to the pseudo-haploid representation that will alternate between the Col-0 and Cvi-0 haplotypes. To assess this, we also aligned a high quality (N50 size = 7.9 Mb) Cvi inbred assembly created with the FALCON assembler[7] to the TAIR10 reference using nucmer using the same parameters as above. From this, we find that 17.3 Mb (14.5%) of the reference is also not found in the Cvi assembly and the Cvi assembly contains 17.7 Mb not found in the reference highlighting the widespread structural variations between the accessions. We also found that the vast majority (94.5%) of the bases from the Pseudohaploid assembly that were not aligned to the reference genome could be successfully aligned to the Cvi assembly using the same parameters. The remainders tended to be of either short or low complexity sequences that are difficult to align using these parameters.

Overall, the Pseudohaploid method was highly effective: it removed 71 Mb of redundant sequences to substantially improve the fraction of unique genes while only marginally decreasing the sequences from the reference present in the pseudohaploid assembly. We have highlighted these results and made these assembly files available on the Pseudohaploid github repository webpage.

Hi-C Sequencing and Scaffolding. Hi-C libraries were created from tender leaves of CB5 in BioMarker Technologies Company as described before[12]. Briefly, the leaves were fixed with formaldehyde, lysed, and then the cross-linked DNA digested with *Hin*d III overnight. Sticky ends were biotinylated and proximity-ligated to form chimeric junctions, that were enriched for and then physically sheared to a size of 500-700 bp. Chimeric fragments representing the original cross-linked long-distance physical interactions were then processed into paired-end sequencing libraries and sequenced on illumina HiSeq X10 platform. The quality of Hi-C experiments was assessed using a list of parameters by HiC-Pro[13] (Supplementary Table 27). The validate rate is 88.22% and dangling end rate is 6.86, indicating a high-quality Hi-C library. To construct chromosome-level assembly, paired-end reads were mapped onto the draft assembly using juicer pipeline with default parameters[14]. Mis-joined contigs with more than 15 kb length were first corrected by detecting abrupt long-range contact patterns using the 3D-DNA pipeline[15]. Subsequently, Hi-C reads were re-mapped to the corrected contigs using bwa aln program and scaffolding was performed using ALLHiC pipeline[16] with the following commands:

```
ALLHiC_partition -b sample.clean.bam -r draft.asm.fasta -e AAGCTT -k 25
allhic extract sample.clean.bam draft.asm.fasta --RE AAGCTT
for K in {1..25};do allhic optimize sample.clean.counts_AAGCTT.25g${K}.txt sample.clean.clm;done
ALLHiC_build draft.asm.fasta
```

Assessment of CB5 genome assembly. To validate the genome consistency, we mapped the illumina short reads to CB5 genome assembly. 98.47% (30.2/30.7 Gb) reads are mappable and covered 99.51% of genome sequences. Additionally, we found 5.2 Mb regions with low coverage (< 5 reads), accounting for only 1.01% of assembled genome (Supplementary Table 4). After mapping the short reads, genome analysis toolkit (GATK) was applied to identify the SNPs with the HaplotypeCaller algorithm. 78,620 homozygous variants were identified in our analysis and accounted for 0.015% of the genome sequences (Supplementary Table 4), indicating high reliability of the consensus sequences. In addition, BUSCO analysis indicates high completeness of our genome assembly, with 92.6% completeness of this genome and only 3.6% of the genome missing (Supplementary Table 2). We further validate the genome assembly by mapping the RNA-seq assembled transcripts (Supplementary Table 3). Our analysis identified 99.88% transcript bases could be covered by the CB5 assembly with high accuracy (99.2%).

CB5 genome annotation. RNA-seq samples were imported into Trinity *de novo* assembly and genome-guided assembly pipelines independently with default parameters[17]. Both of the assembly strategies were based on whole RNA-seq datasets. RSEM was used to calculate transcript abundance[18]. Transcripts with FPKM < 1 and iso-percentage < 3 % were removed from further analysis. The filtered transcripts were imported to PASA program for construction of comprehensive transcripts by integrating transcripts generated from Trinity *de novo* and genome-guided pipelines as PASA is able to take advantage of the high sensitivity of reference-based assembly while leveraging the ability of *de novo* assembly to detect novel transcripts[19]. The PASA-assembled transcripts described above were used for training. The nearly "full-length" transcripts were evaluated by comparing with UniProt plant protein database (last accessed on 8 December 2016) and proteins that were covered at least 95 % were retained as candidates. Then *ab initio* gene predictors, including SNAP[20], GENEMARK[21] and AUGUSTUS[22], were each trained with those selected proteins. After that, MAKER pipeline was used to integrate multiple tiers of coding evidence, including *ab initio* gene prediction, transcript evidence and protein evidence, and generate a comprehensive set of protein-coding genes. The final release of CB5 annotation includes a total of 29,415 protein-coding genes. BUSCO assessment reported 84.7% of completeness (Supplementary Table 28).

Repeat prediction. We first customized a *de novo* repeat library of the genome using RepeatModeler (http://www.repeatmasker.org/RepeatModeler/), which can automatically execute two *de novo* repeat finding programs, including RECON (version 1.08)[23] and RepeatScout (version 1.0.5)[24]. The consensus TE sequences generated above were imported to RepeatMasker (version 4.05)[25] to identify and cluster repetive elements. Unknown TEs were further classified using TEclass (version 2.1.3)[26]. To identify tandem repeats within the genome, the Tandem Repeat Finder (TRF) package (version 4.07)[27] was used with the modified parameters of "1 1 2 80 5 200 2000 –d -h" in order to find high order repeats. Telomeres and centromeres were identified based on the .dat output files above. Repeat sequences with more than 10 monomers 'AAACCT' were identified as telomeres. For centromeres identification, we used a similar method described in *Oropetium thomaeum* genome[28]. To calculate the insertion time, we referred to Kimura distance to indicate the recent TE copies and old TE copies (Kimura value ranged from 0 to 50). A substitution rate of 1.38×10^{-8} per site per year was used and insertion time was estimated as $T = K/2\mu$ (K is the divergence rate and μ is the neutral mutation rate).

References

1 Ming, R. *et al.* Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* **14**, R41, (2013).

2 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736 (2017).

3 Ming, R. *et al.* The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435-1442 (2015).

4 Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130-135 (2012).

5 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).

6 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).

7 Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050-1054 (2016).

8 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757-767 (2017).

9 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

10 Kent, W. J., 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. U S A* **100**, 11484-11489 (2003).

11 Nattestad, M. & Schatz, M. C. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021-3023 (2016).

12 Xie, T. *et al.* De novo plant genome assembly based on chromatin interactions: a case study of Arabidopsis thaliana. *Mol. Plant* **8**, 489-492 (2015).

13 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).

14 Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* **3**, 95-98 (2016).

15 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).

16 Zhang, J. *et al.* Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **50**, 1565-1573 (2018).

17 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512, (2013).

18 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

19 Brian J. Haas *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **21**, 5654-5666 (2003).

20 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, (2004).

21 Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O. & Borodovsky, M. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* **33**, 6494-6506 (2005).

22 Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).

23 Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269-1276 (2002).

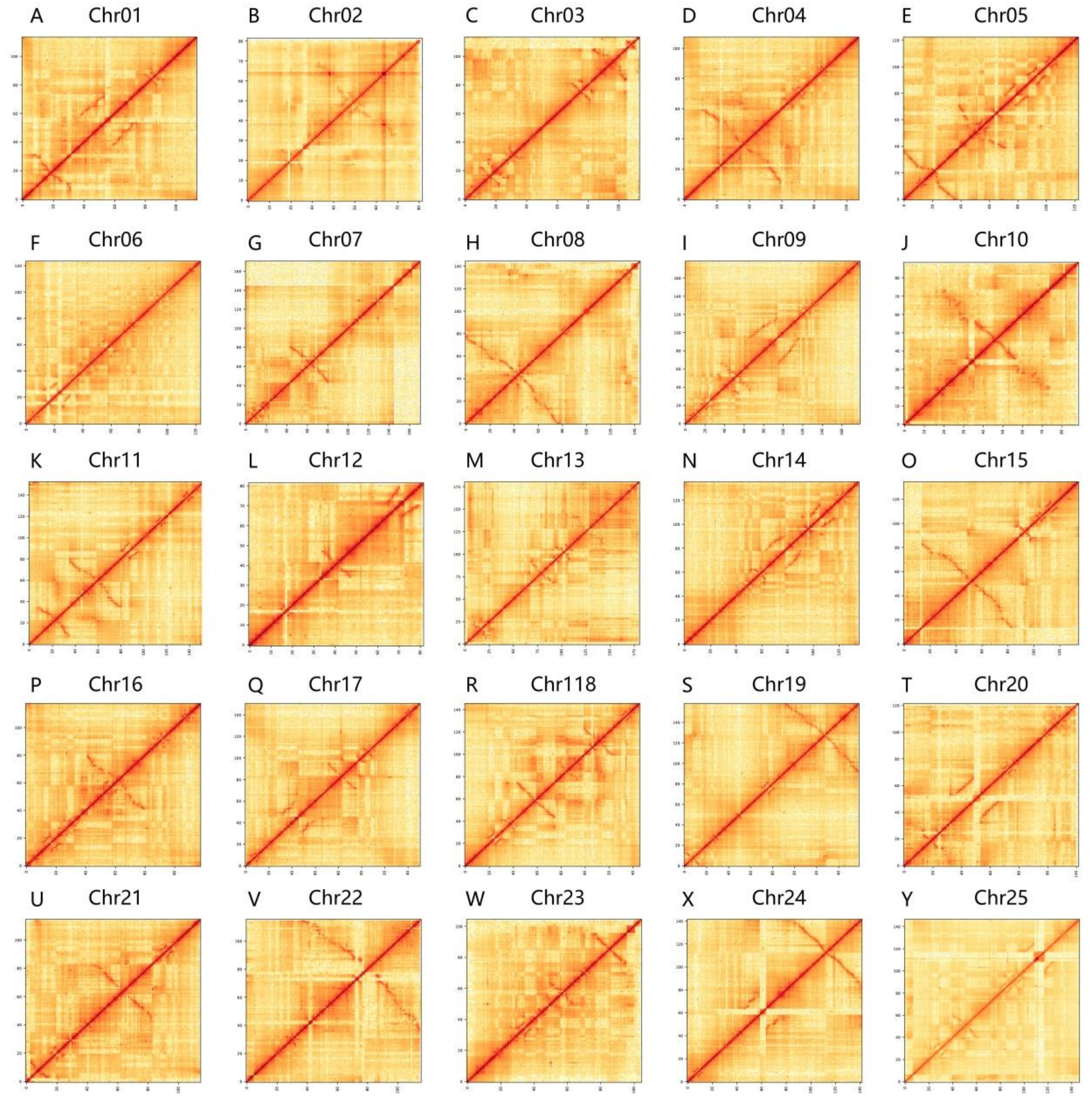
24 Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-358 (2005).

25 Smit, A., Hubley, R & Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org. (2013-2015).

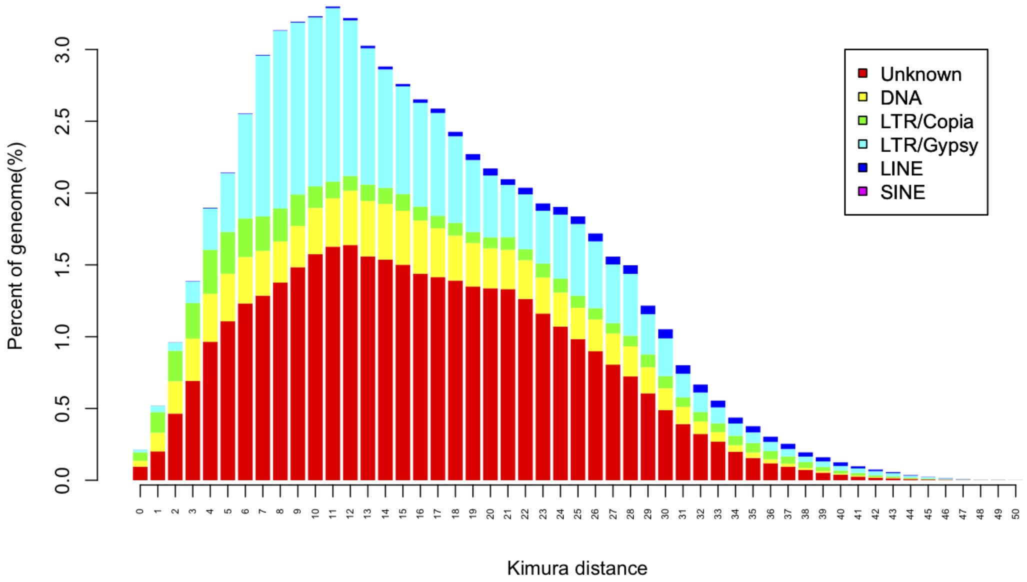
26 Abrusan, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass--a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329-1330 (2009).

27 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573-580 (1999).

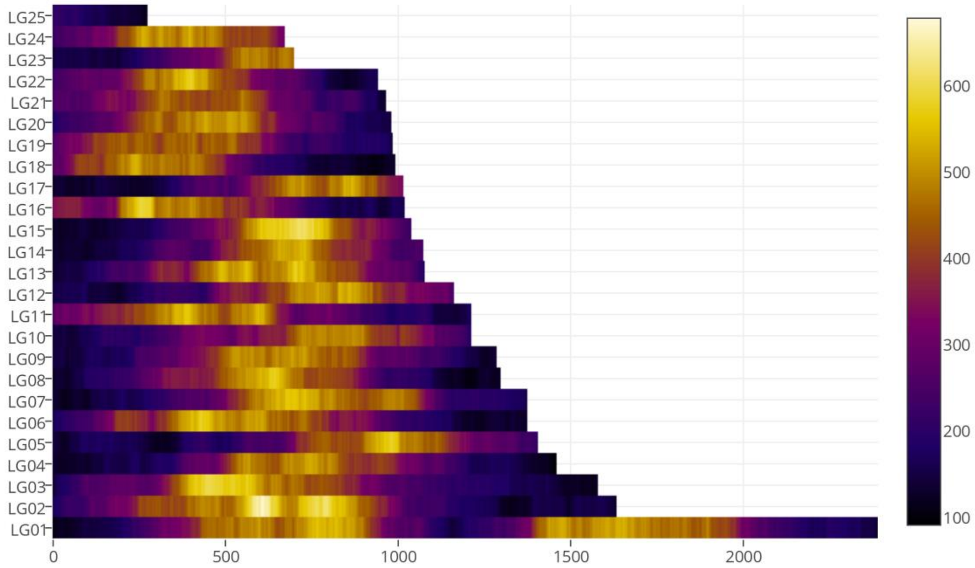
28 VanBuren, R. *et al.* Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* **527**, 508-511 (2015).



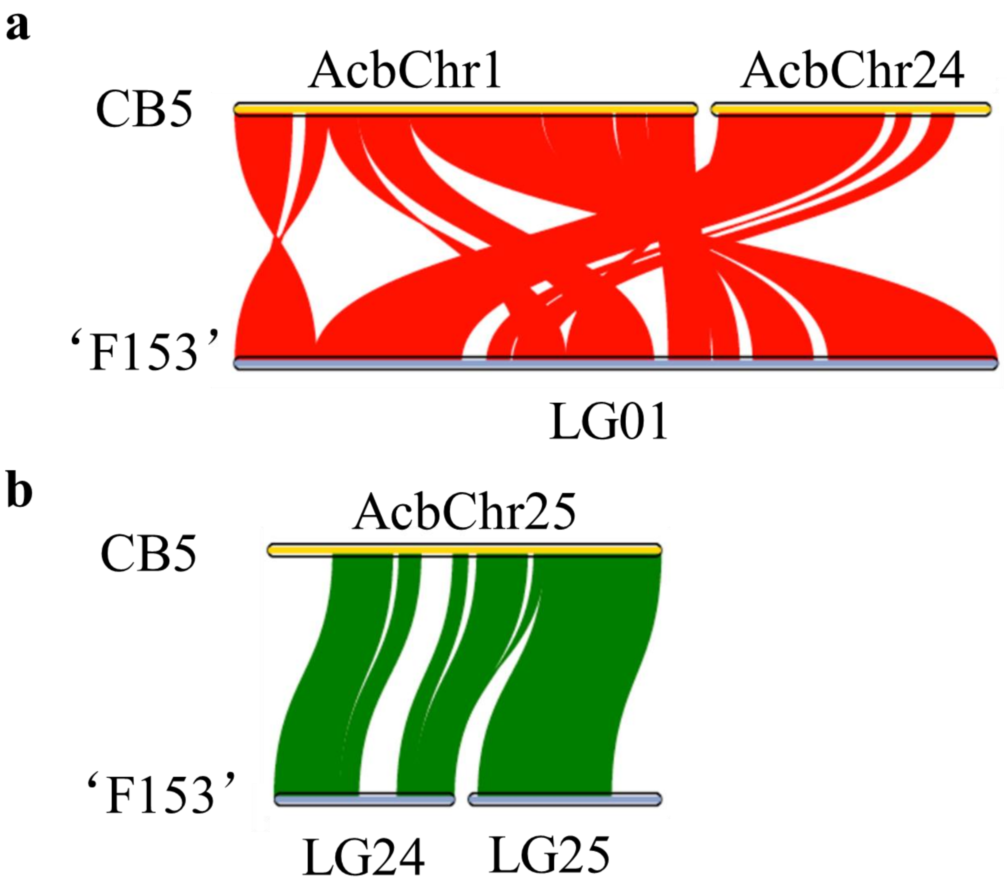
Supplementary Figure 1. Genome-wide analysis of chromatin interactions at 150-kb resolution in pineapple CB5 genome.



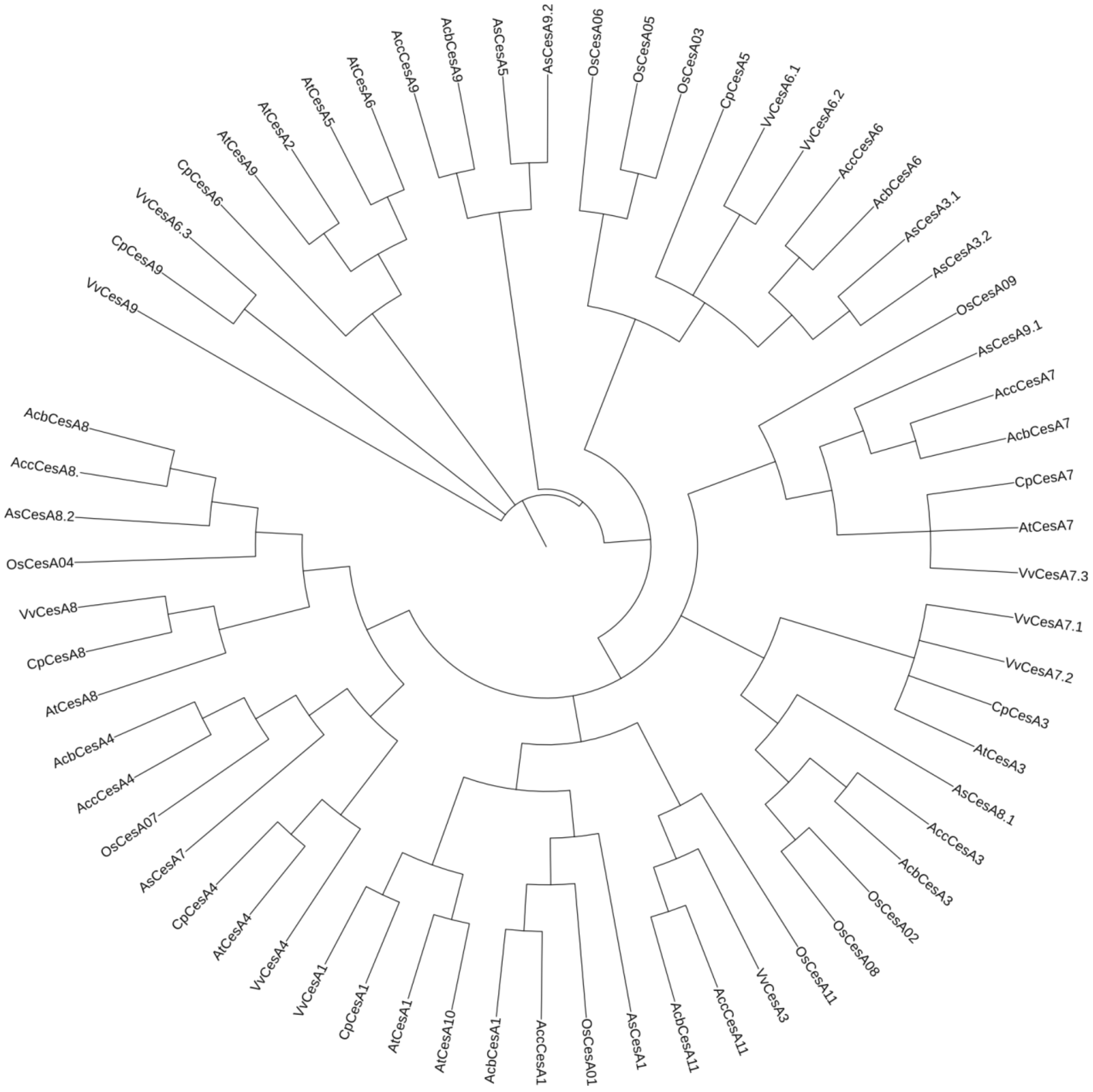
Supplementary Figure 2. Kimura distance-based copy divergence analysis of transposable elements in CB5 genome. The graph represents percentage of genome (y-axis) of each type of TEs (SINE, LINE, LTR/Gypsy, LTR/Copia and DNA transposons), clustered according to Kimura distances to their corresponding consensus sequences (x-axis, K-value from 0 to 50).



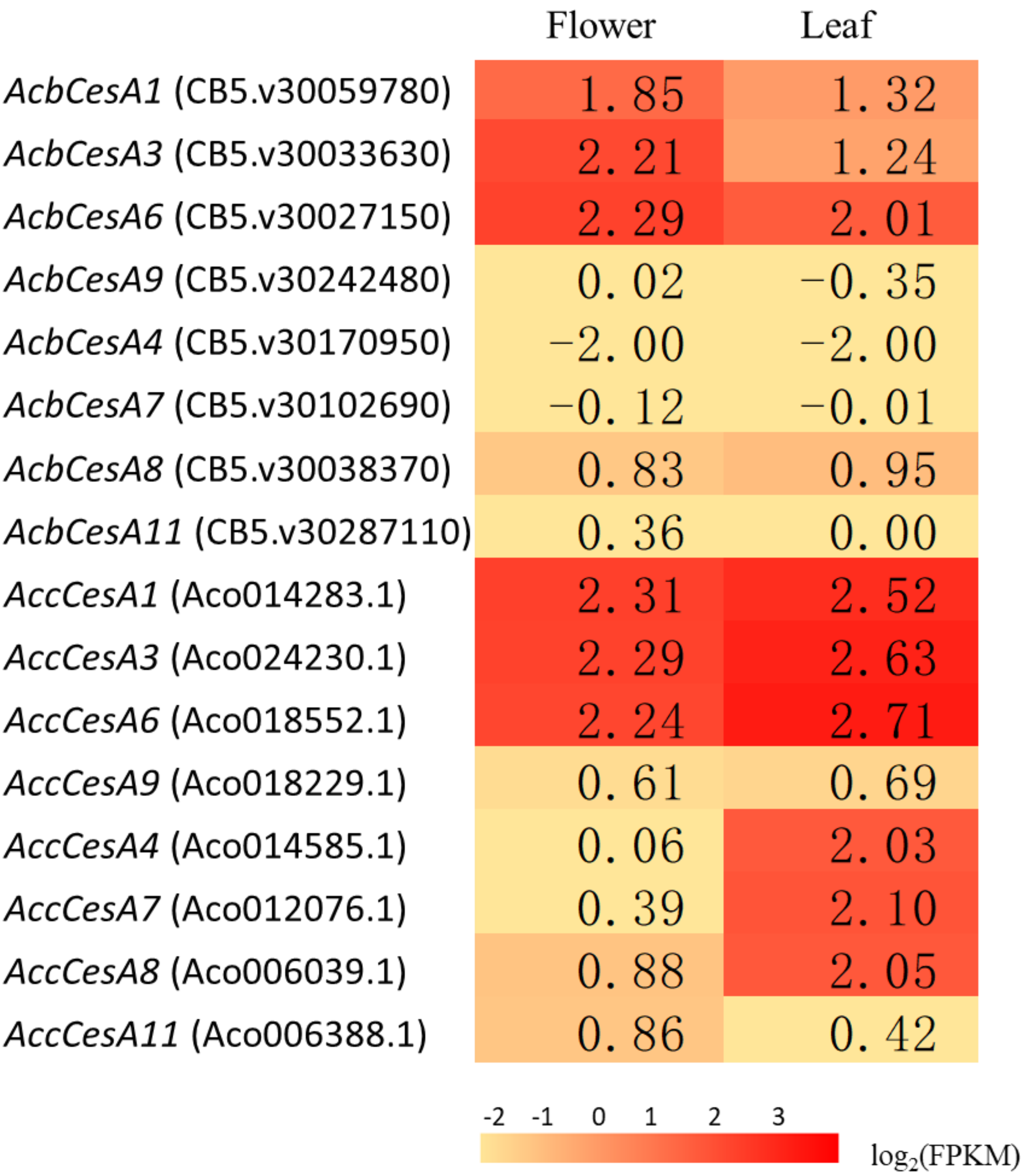
Supplementary Figure 3. Gypsy density of 'F153' genome.



Supplementary Figure 4. Synteny analysis between CB5 and ‘F153’. a, Synteny between LG01 of F153 and AcbChr1 and AcbChr24 of CB5. **b**, Synteny between LG24 and LG25 of F153 and AcbChr25 of CB5.

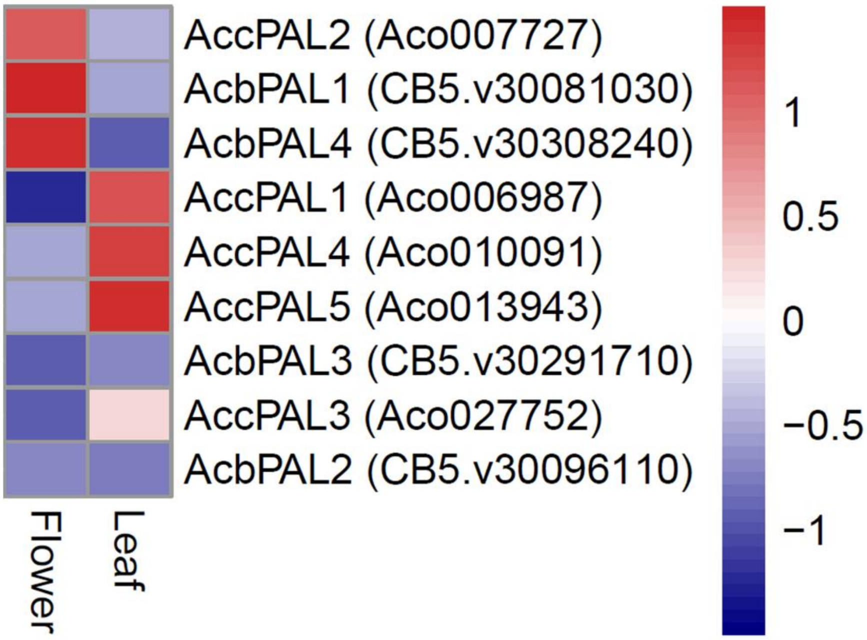


Supplementary Figure 5. Evolutionary relationships of CesA proteins. The evolutionary history was inferred using the Neighbor-Joining method. CesA homologue sequences obtained from NCBI for the species *Arabidopsis thaliana* (*At*), *Oryza sativa* (*Os*), *Carica papaya* (*Cp*), *Vitis vinifera* (*Vv*) and *Apostasia shenzhenica* (*As*).The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 64 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 139 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

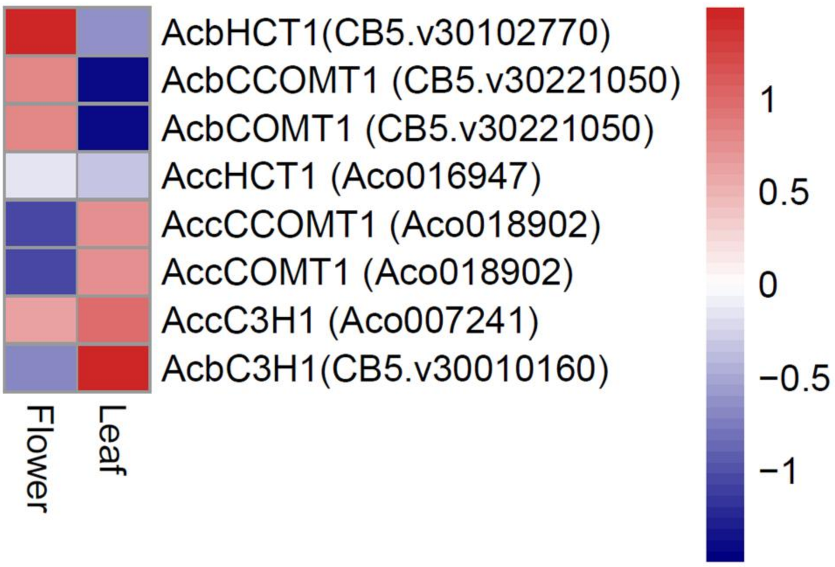


	Flower	Leaf
AcbCesA1 (CB5.v30059780)	1. 85	1. 32
AcbCesA3 (CB5.v30033630)	2. 21	1. 24
AcbCesA6 (CB5.v30027150)	2. 29	2. 01
AcbCesA9 (CB5.v30242480)	0. 02	−0. 35
AcbCesA4 (CB5.v30170950)	−2. 00	−2. 00
AcbCesA7 (CB5.v30102690)	−0. 12	−0. 01
AcbCesA8 (CB5.v30038370)	0. 83	0. 95
AcbCesA11 (CB5.v30287110)	0. 36	0. 00
AccCesA1 (Aco014283.1)	2. 31	2. 52
AccCesA3 (Aco024230.1)	2. 29	2. 63
AccCesA6 (Aco018552.1)	2. 24	2. 71
AccCesA9 (Aco018229.1)	0. 61	0. 69
AccCesA4 (Aco014585.1)	0. 06	2. 03
AccCesA7 (Aco012076.1)	0. 39	2. 10
AccCesA8 (Aco006039.1)	0. 88	2. 05
AccCesA11 (Aco006388.1)	0. 86	0. 42

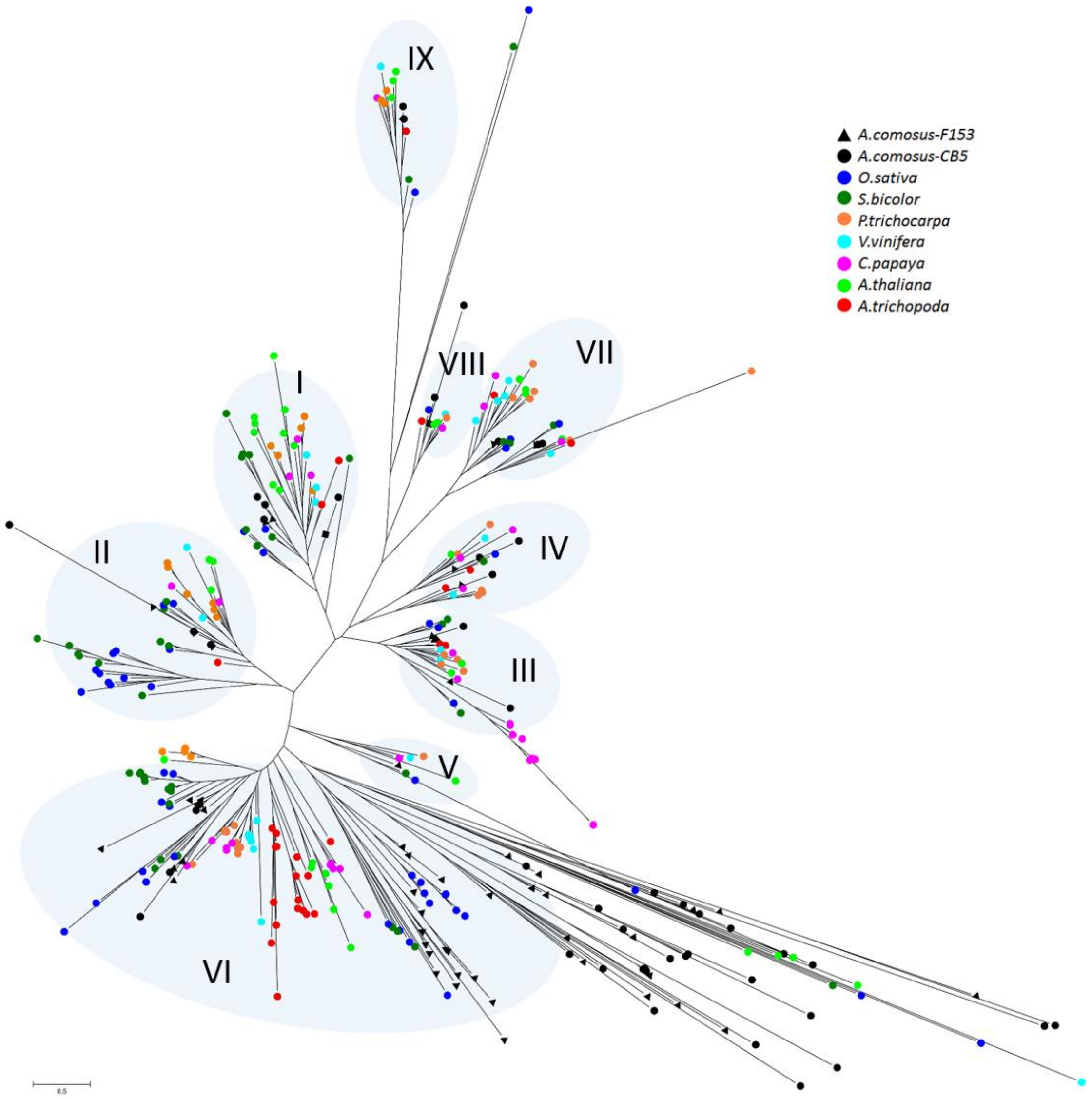
Supplementary Figure 6. The cellulose synthesis genes' expression (Log$_2$ of FPKM values) pattern in CB5 and 'F153'.



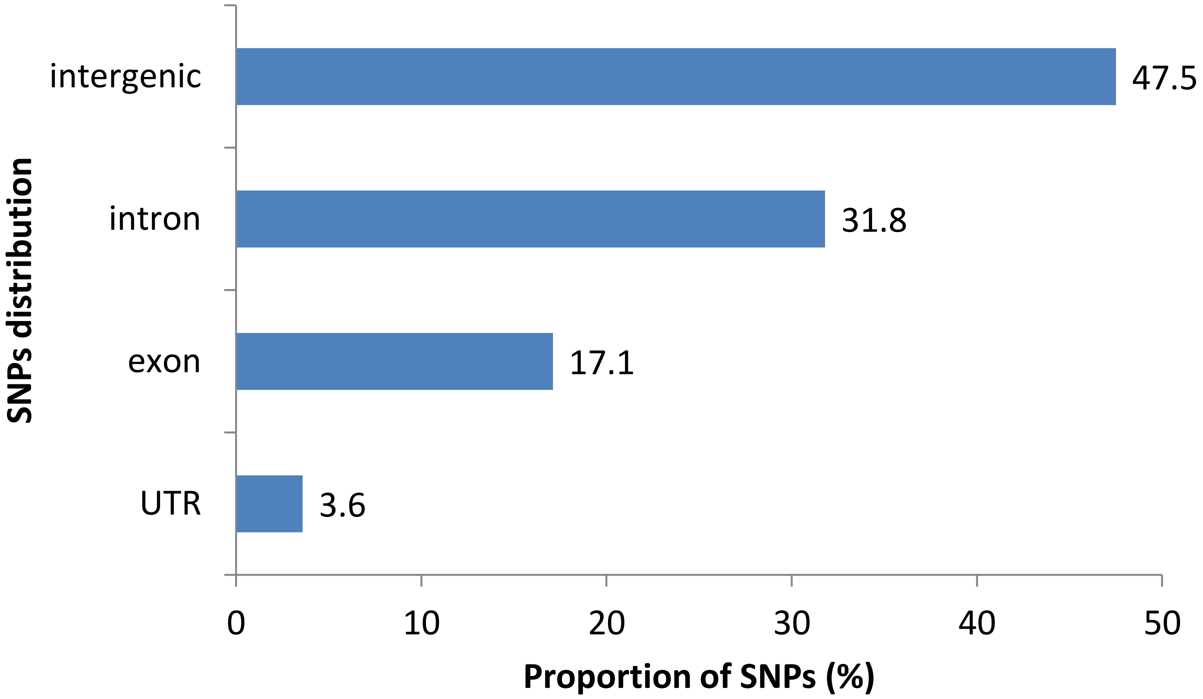
Supplementary Figure 7. Relative expression of *PAL* genes in 'F153' and CB5 in different tissues. *PAL* genes Log_2 transformed FPKMs of expression in flower, immature fruit, ripen fruit and leaf issue.



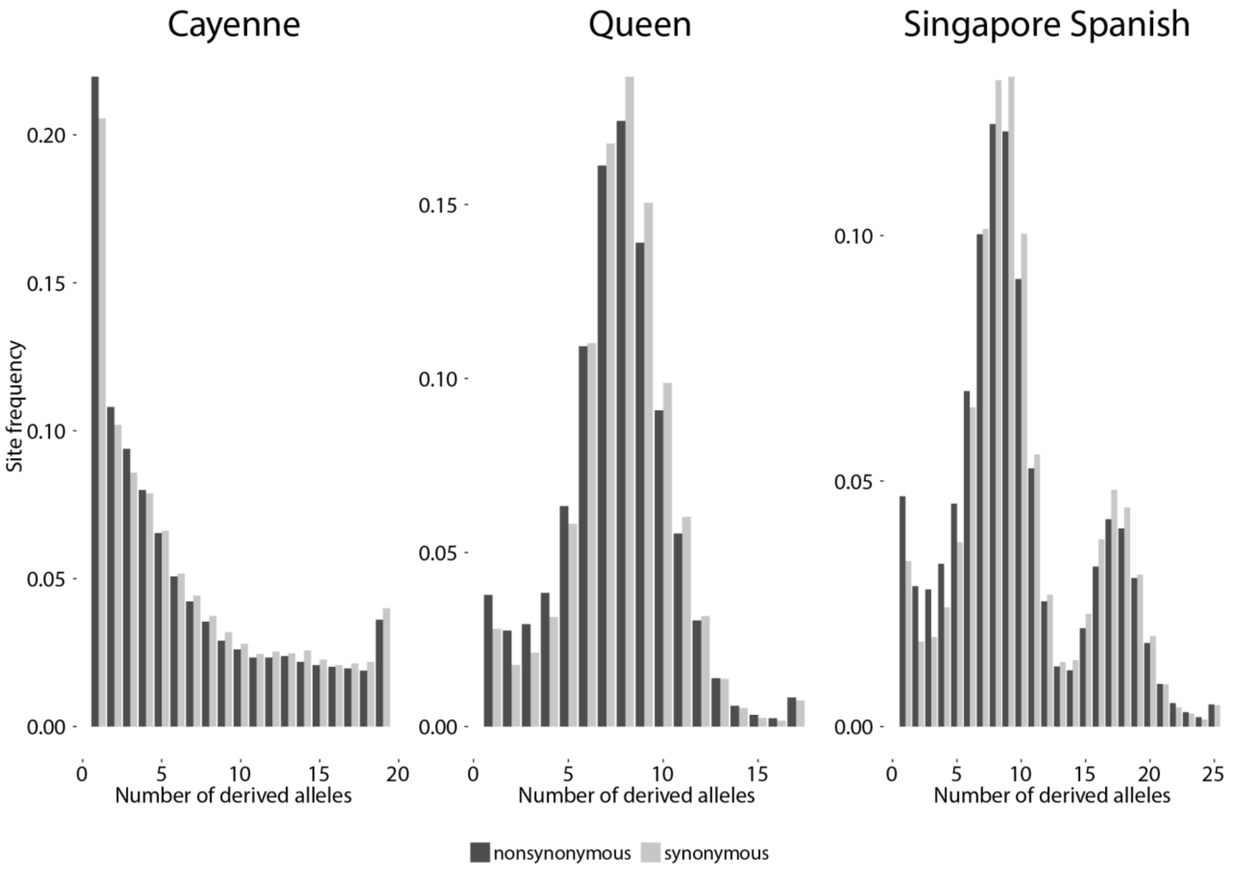
Supplementary Figure 8. Relative expression of *HCT*, *CCOMT* and *COMT* genes in 'F153' and CB5 in different tissues. *HCT, CCOMT* and *COMT* genes Log_2 transformed FPKMs of expression in flower, immature fruit, ripen fruit and leaf issue.



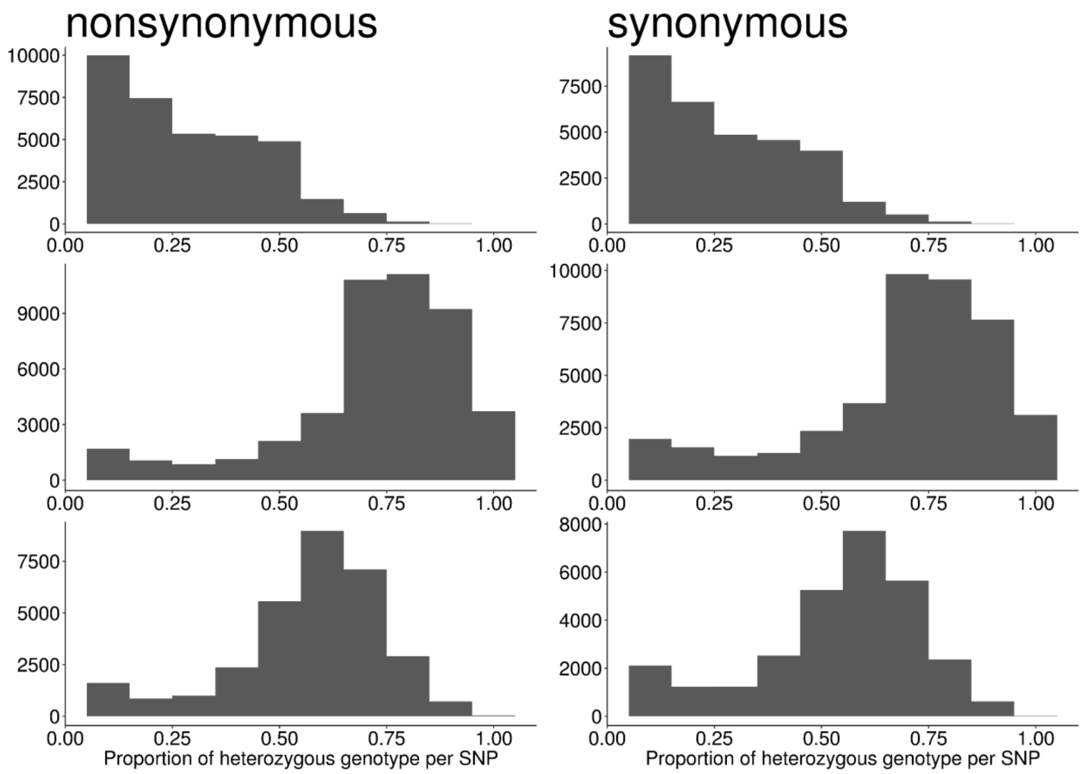
Supplementary Figure 9. The phylogenetic tree of 378 CP genes identified in 'F153', CB5, *Amborella*, *Populus*, *Arabidopsis*, grape, papaya, sorghum and rice genomes.



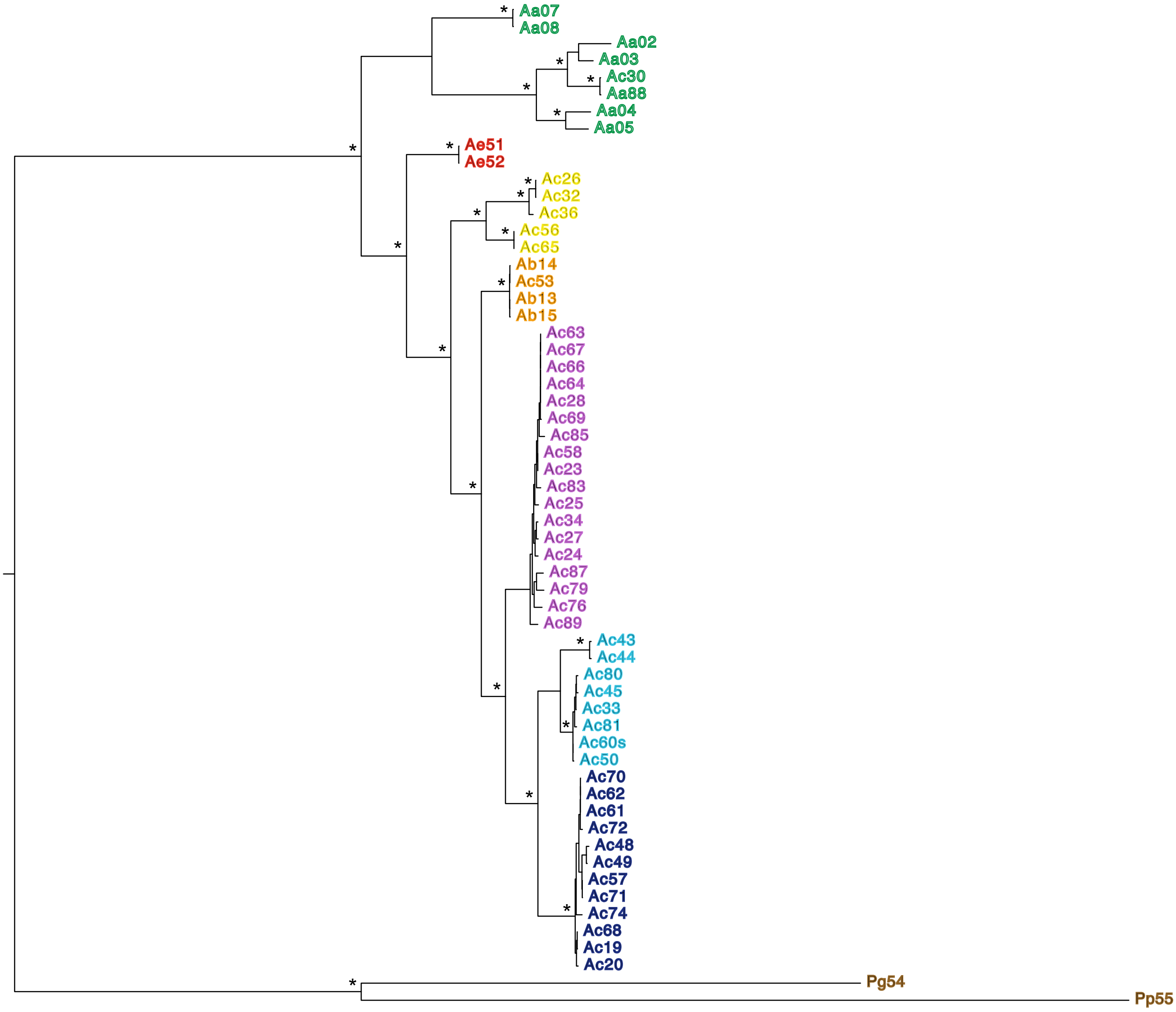
Supplementary Figure 10. Proportion of SNPs in different regions across the pineapple 'F153' genome.



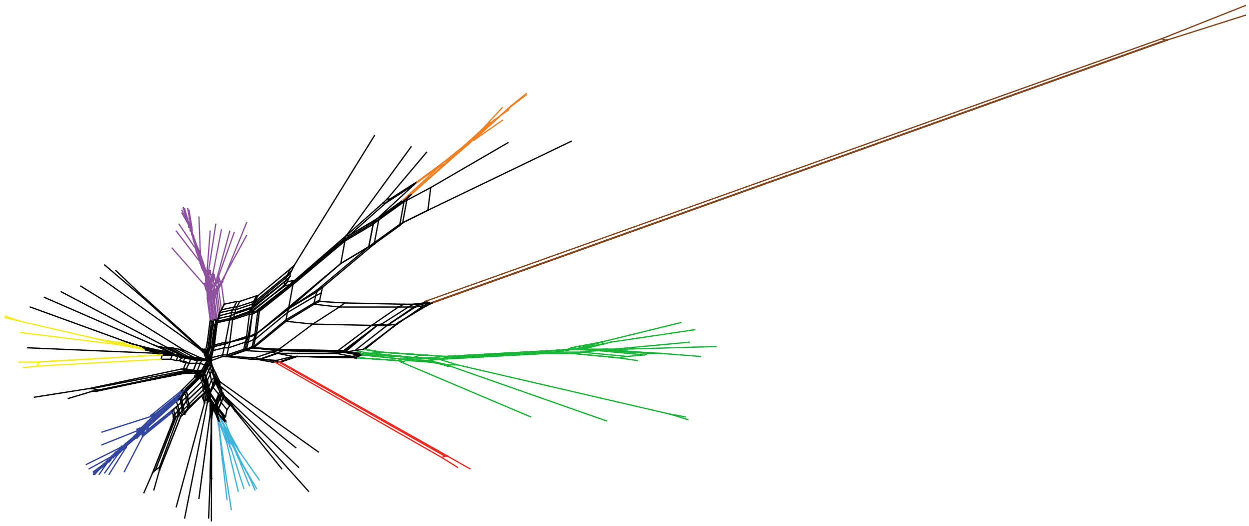
Supplementary Figure 11. **Nonsynonymous and synonymous site frequency spectrum for cultivars Smooth Cayenne, Queen, and Singapore Spanish.**



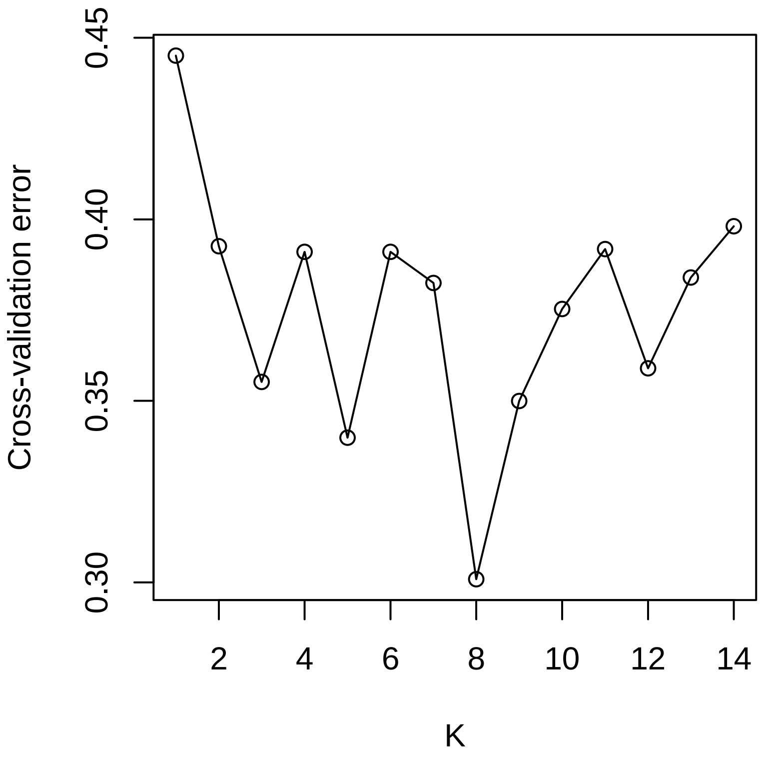
Supplementary Figure 12. Histogram of proportion of heterozygote genotype per SNP position. Top row represents cultivar Smooth Cayenne, middle row represents cultivar Queen, and the bottom row represents cultivar Singapore Spanish.



Supplementary Figure 13. Maximum likelihood (RAxML) tree of all non-admixed *Ananas* accessions. Asterisks indicate branches with 100% bootstrap support. Green, variety *microstachys*; red, variety *erectifolius*; orange, variety *bracteatus*; yellow, variety *comosus* / Mordilona-related cultivar cluster; violet, variety *comosus* / cultivar Singapore Spanish culster; light blue, variety *comosus* / cultivar Smooth Cayenne cluster; dark blue, variety *comosus* / cultivar Queen cluster; brown, genus *Pitcairnia* outgroups.



Supplementary Figure 14. SplitsTree network for all studied *Ananas* accessions including admixed samples.

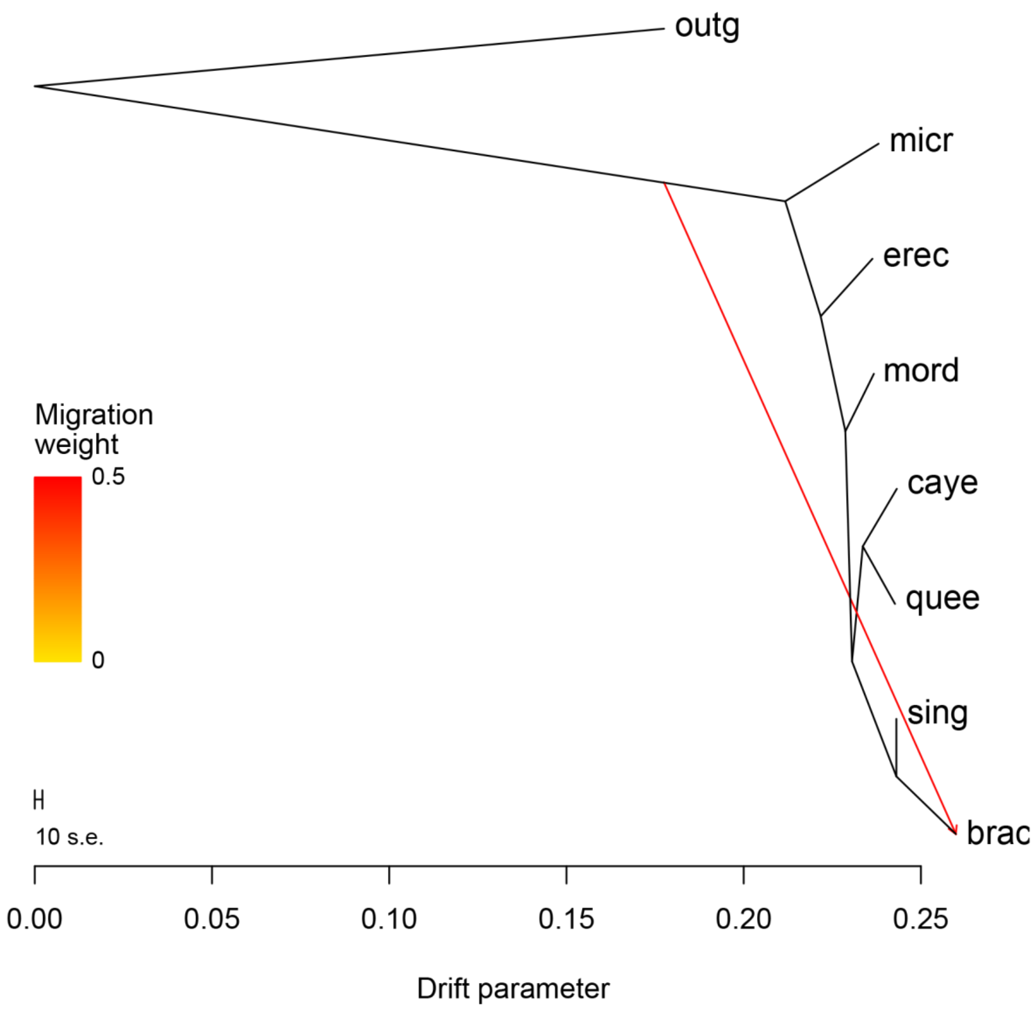


Supplementary Figure 15. Cross-validation error of ADMIXTURE analysis.

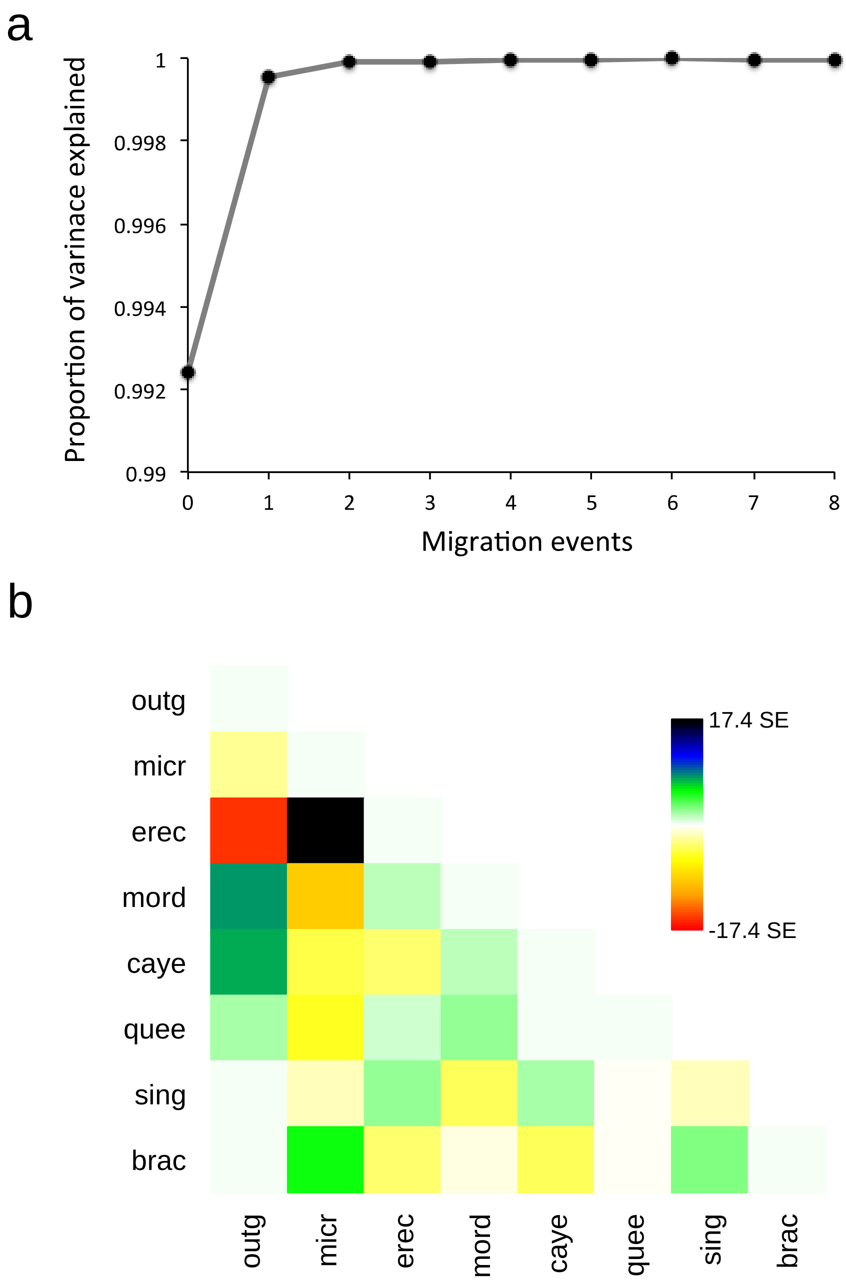
	micr	erec	brac	como	mord	caye	quee	sing
micr	0.0047	0.70	0.64	0.62	0.61	0.71	0.73	0.72
erec	0.0083	0.0003	0.82	0.73	0.71	0.89	0.94	0.93
brac	0.0083	0.0046	0.0002	0.39	0.51	0.66	0.74	0.64
como	0.0091	0.0046	0.0030	0.0022	0.28	0.35	0.39	0.33
mord	0.0092	0.0047	0.0039	0.0032	0.0024	0.52	0.60	0.61
caye	0.0094	0.0045	0.0031	0.0022	0.0033	0.0007	0.78	0.81
quee	0.0091	0.0046	0.0033	0.0020	0.0034	0.0021	0.0003	0.90
sing	0.0090	0.0045	0.0024	0.0019	0.0035	0.0026	0.0028	0.0003

Divergence	Diversity	Fst
0.0095	0.0050	0.95
0.009	0.0046	0.9
0.0085	0.0042	0.85
0.008	0.0038	0.8
0.0075	0.0034	0.75
0.007	0.0030	0.7
0.0065	0.0026	0.65
0.006	0.0022	0.6
0.0055	0.0018	0.55
0.005	0.0014	0.5
0.0045	0.0010	0.45
0.004	0.0006	0.4
0.0035	0.0002	0.35
0.003		0.3
0.0025		0.25
0.002		

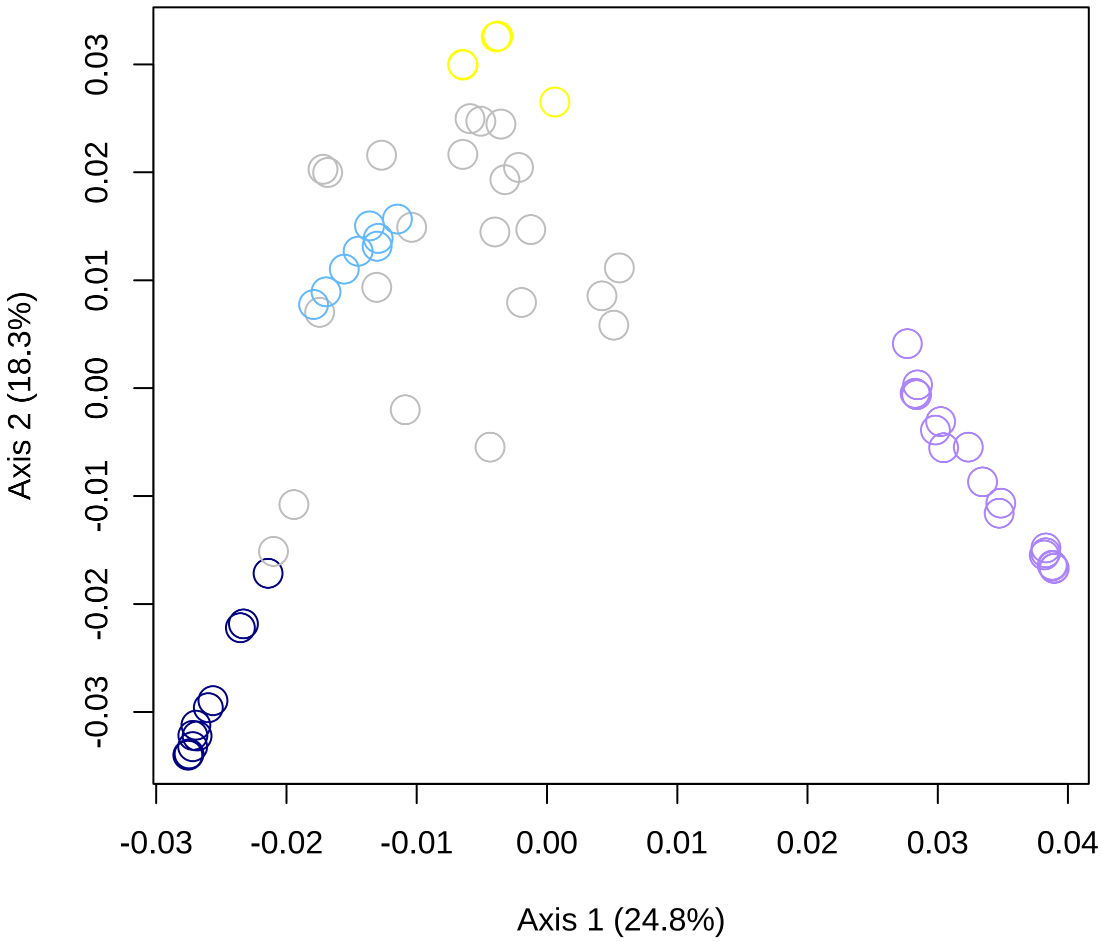
Supplementary Figure 16. Heat map showing nucleotide divergence (Dxy; lower triangle), nucleotide diversity (pi; diagonal) and fixation Index (F_{ST}; upper triangle) among *Ananas* accessions. Micr, variety *microstachys*; erec, variety *erectifolius*; brac, variety *bracteatus*; como, all accessions of variety *comosus*; mord, caye, quee, sing stand for cultivars Mordilona-related, Cayenne, Queen and Singapore Spanish of variety *comosus*.



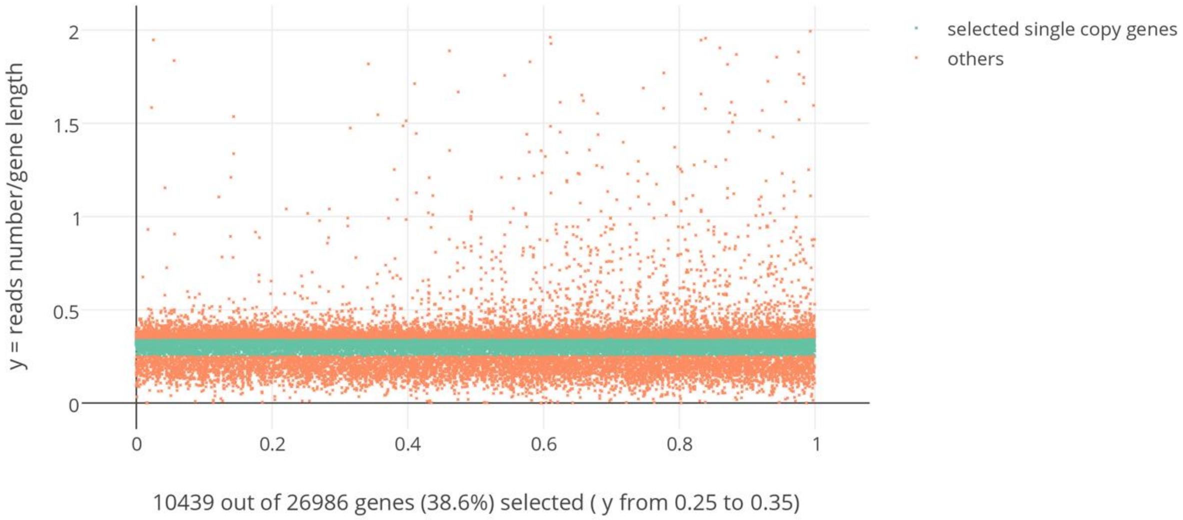
Supplementary Figure 17. TreeMix results. Population graph for the most likely evolutionary scenario, m = 1 migration event.



Supplementary Figure 18. TreeMix results. a, Proportion of variance in relatedness between populations explained by different models with m = 0 to 8. **b**, Residual fit for M = 1 model.

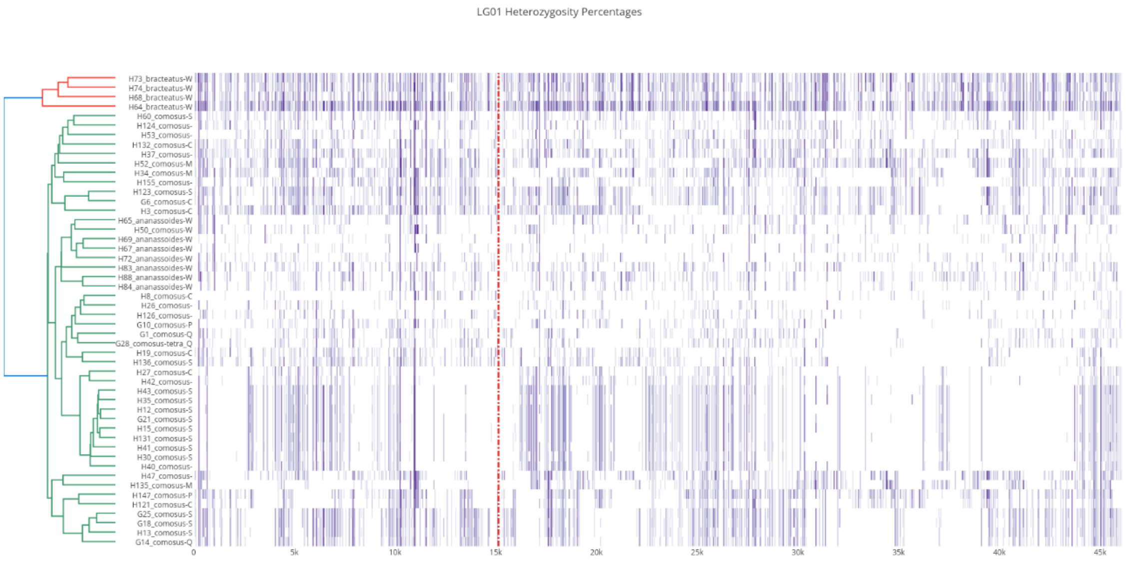


Supplementary Figure 19. Multi dimensional scaling (MDS) graphs of the studied Ananas accessions. Only *comosus* cultivars are shown. Percentage of variance explained by each axis is indicated in bracket. Color code follows Fig. 2, with grey denoting admixed samples (99% criteria).

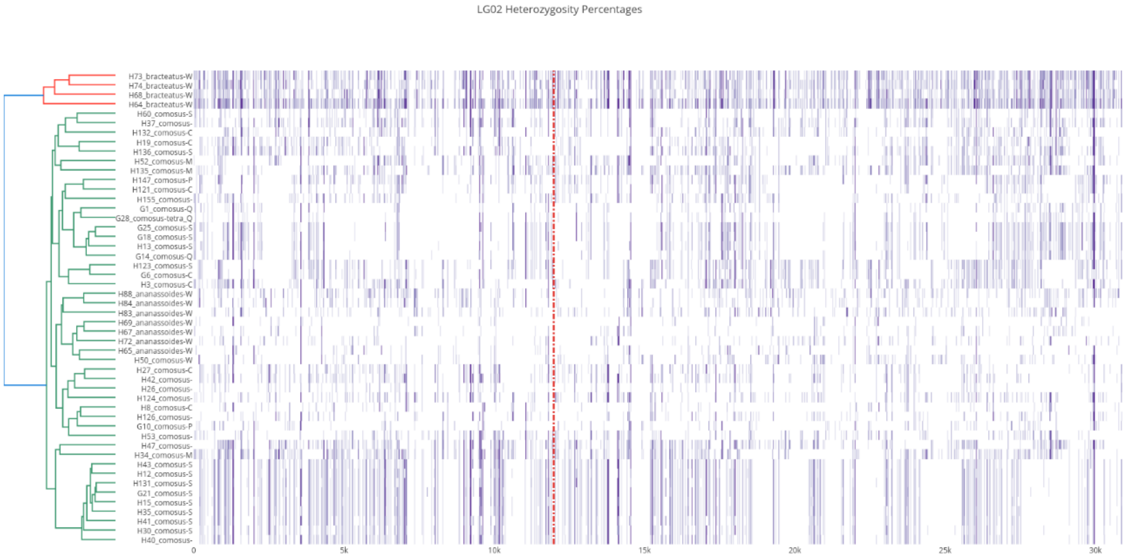


Supplementary Figure 20. Distribution of single copy genes used to identify runs of homozygosity. The normalized number of reads per genes is plotted for the 10,439 single copy genes.

Supplementary Figure 21.1-21.25. Levels of heterozygosity are plotted for every 100 bp across linearly concatenated genes, across 50 varieties and 25 linkage groups. In these heat maps, a higher level of heterozygosity is shown by higher levels of purple. No pigment indicates homozygosity. On each figure, the vertical dotted red line indicates predicted centromere region as determined by *gypsy* LTR retrotransposon abundance. The dendrograms on the left indicate clustered heterozygosity landscapes among varieties. The accession names are formatted as USDA names, variety names below *Ananas comosus* and major cultivar groups delimited by dash. The 5 major cultivar groups ‘Cayenne’, ‘Spanish’, ‘Queen’, and ‘Mordilona-related’ are abbreviated as –C, -S, -Q, and -P respectively while –W indicates wild varieties. As an example, H27-comosus-C is a Cayenne cultivar belonging to *Ananas comosus* var *comosus*, with H27 as the USDA name. Some varieties lack cultivar group information.

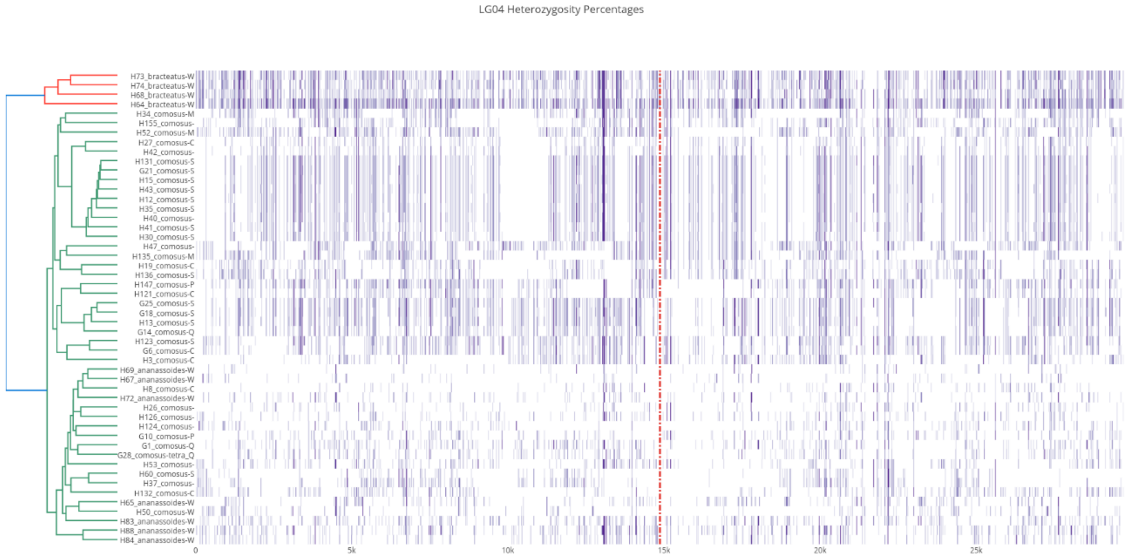


Supplementary Figure 21.1

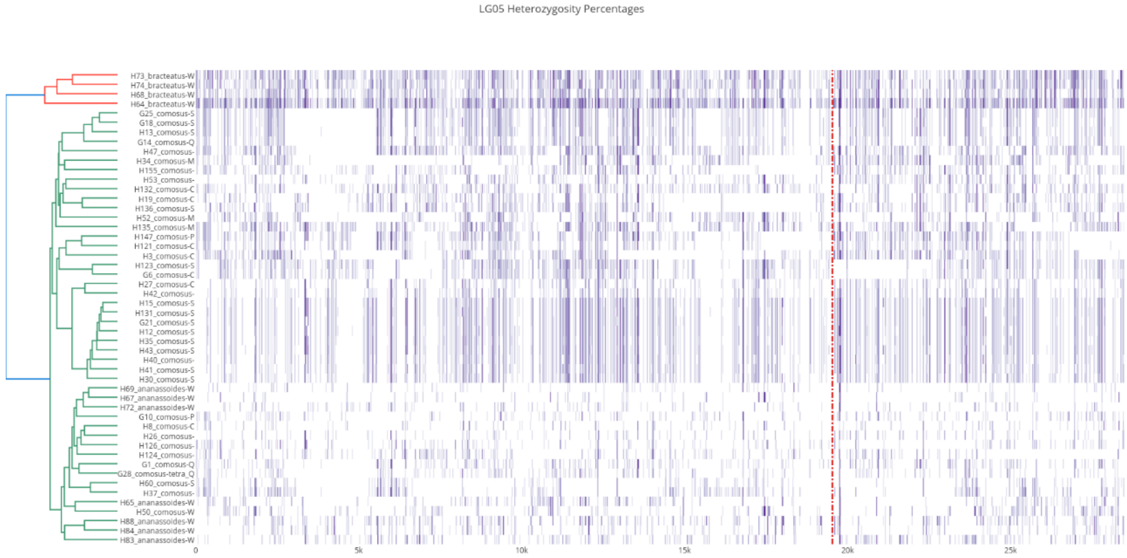


Supplementary Figure 21.2

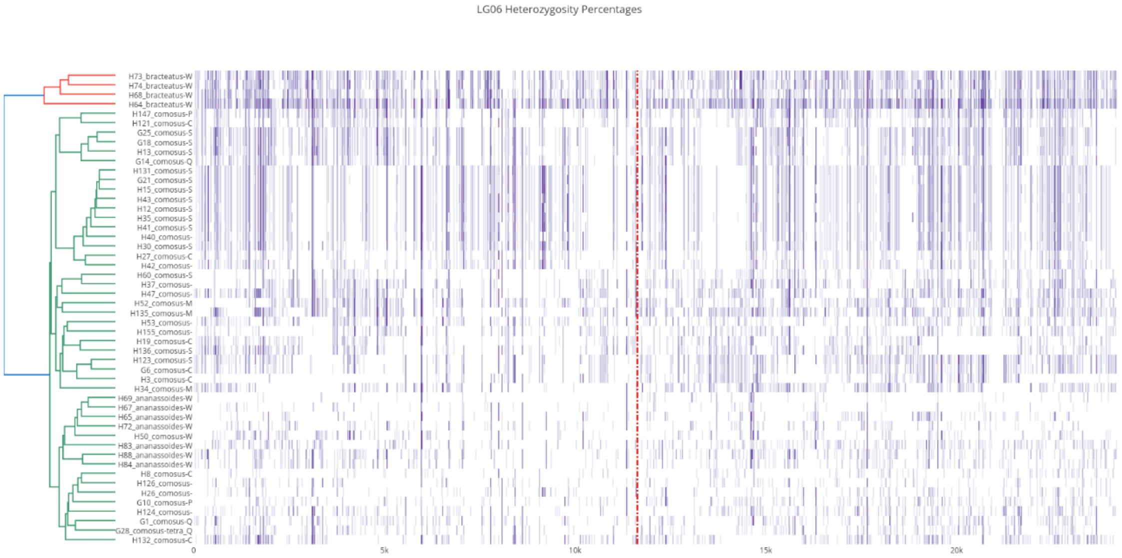
Supplementary Figure 21.3



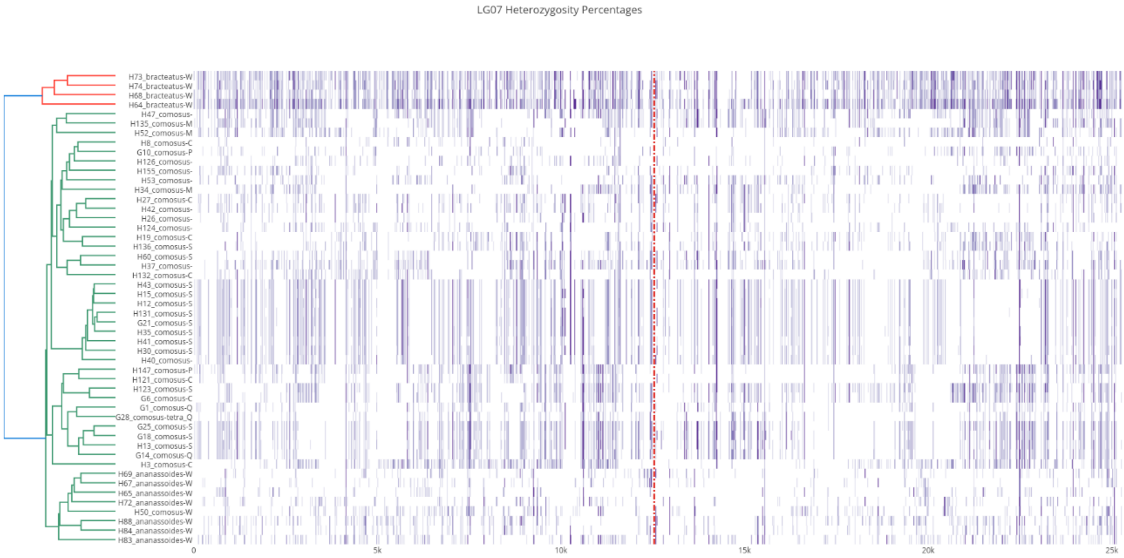
Supplementary Figure 21.4



Supplementary Figure 21.5



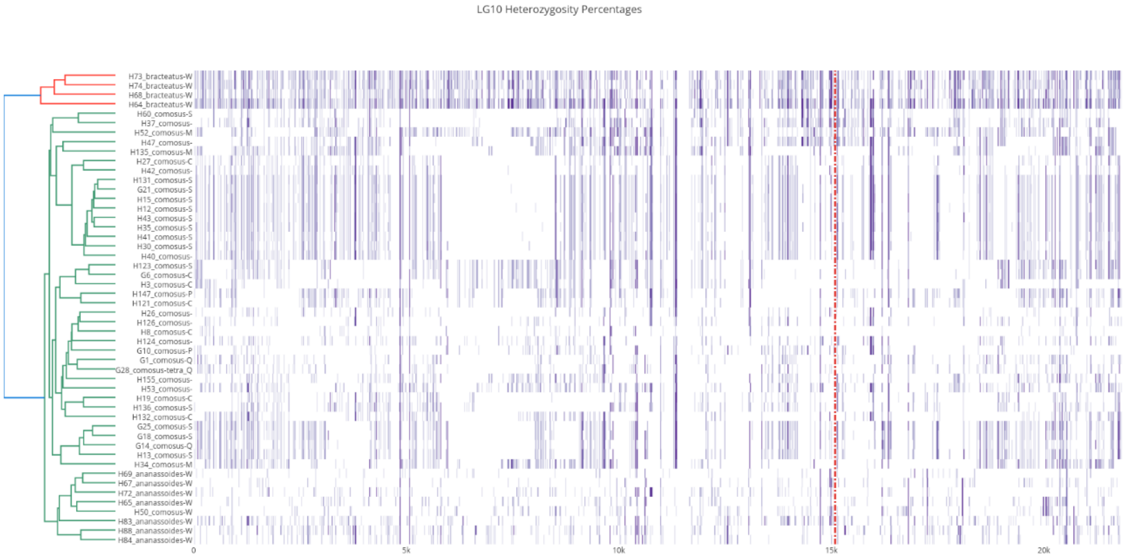
Supplementary Figure 21.6



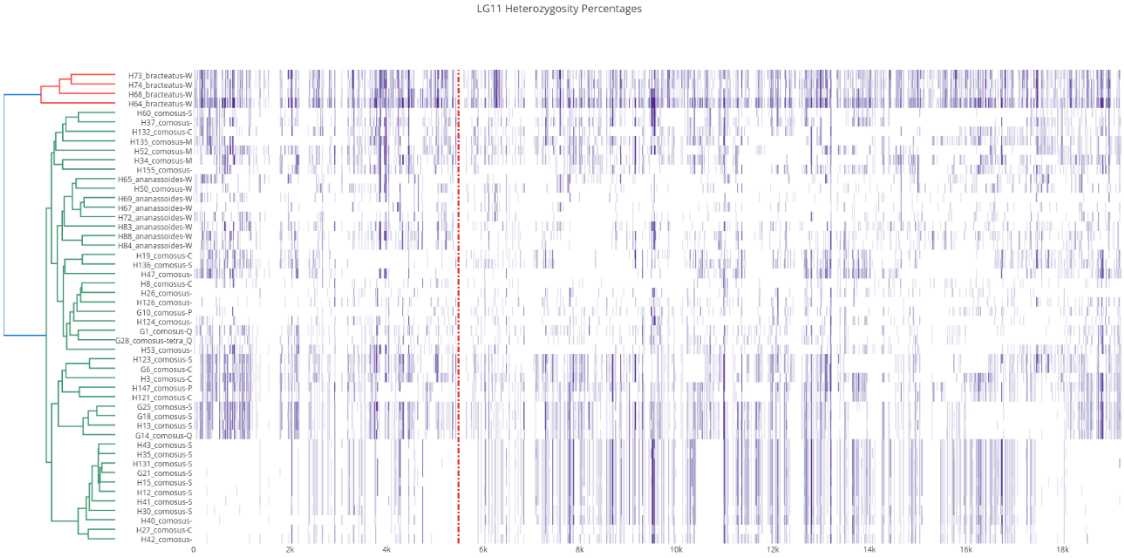
Supplementary Figure 21.7

Supplementary Figure 21.8

Supplementary Figure 21.9



Supplementary Figure 21.10



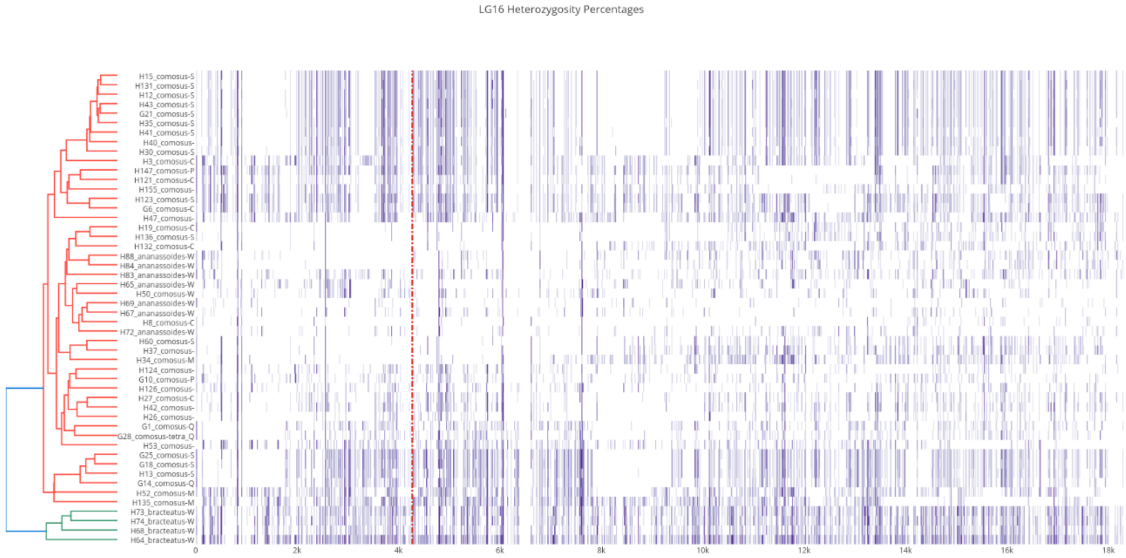
Supplementary Figure 21.11

Supplementary Figure 21.12

Supplementary Figure 21.13

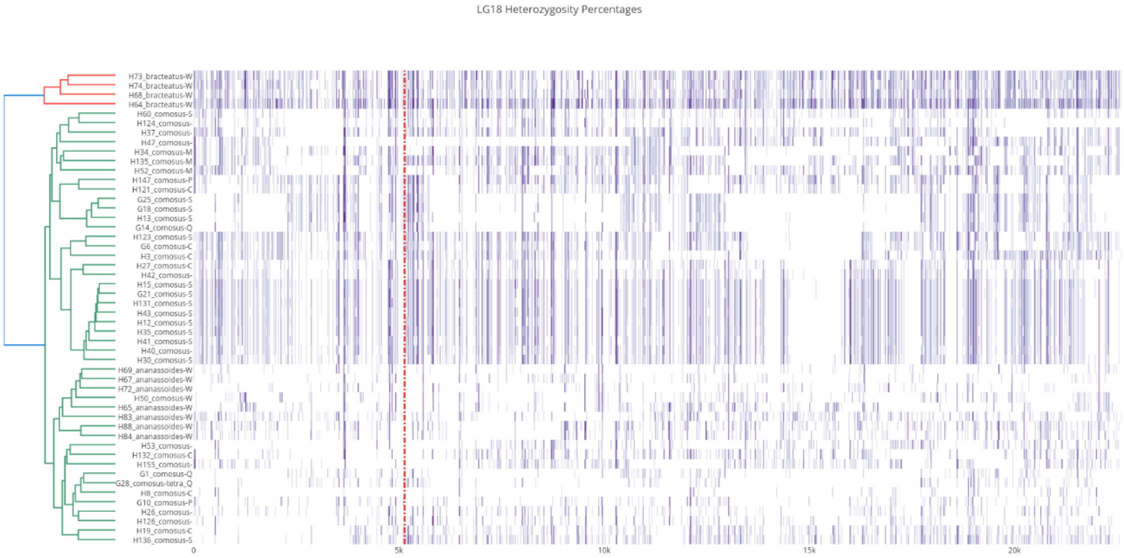
Supplementary Figure 21.14

Supplementary Figure 21.15



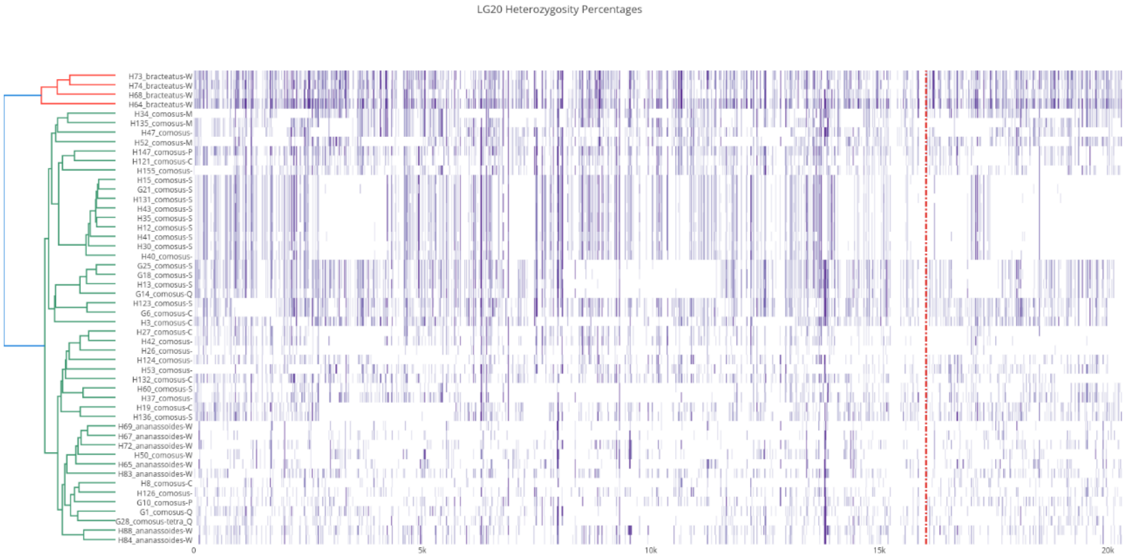
Supplementary Figure 21.16

Supplementary Figure 21.17



Supplementary Figure 21.18

Supplementary Figure 21.19



Supplementary Figure 21.20

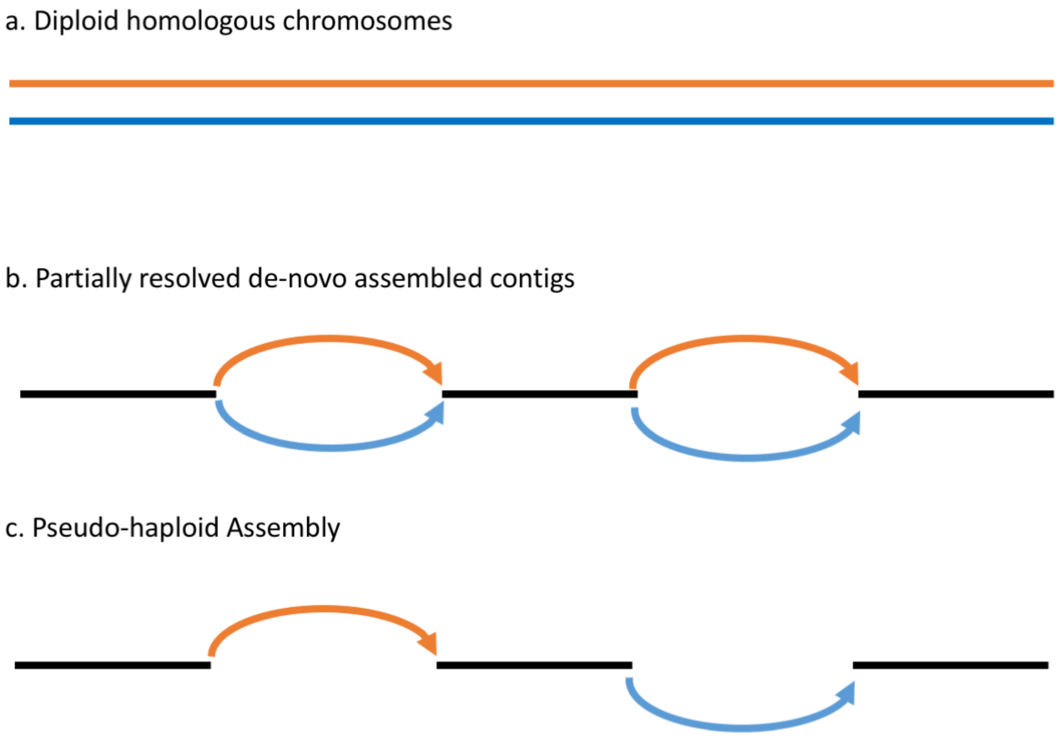
Supplementary Figure 21.21

Supplementary Figure 21.22

Supplementary Figure 21.23

Supplementary Figure 21.24

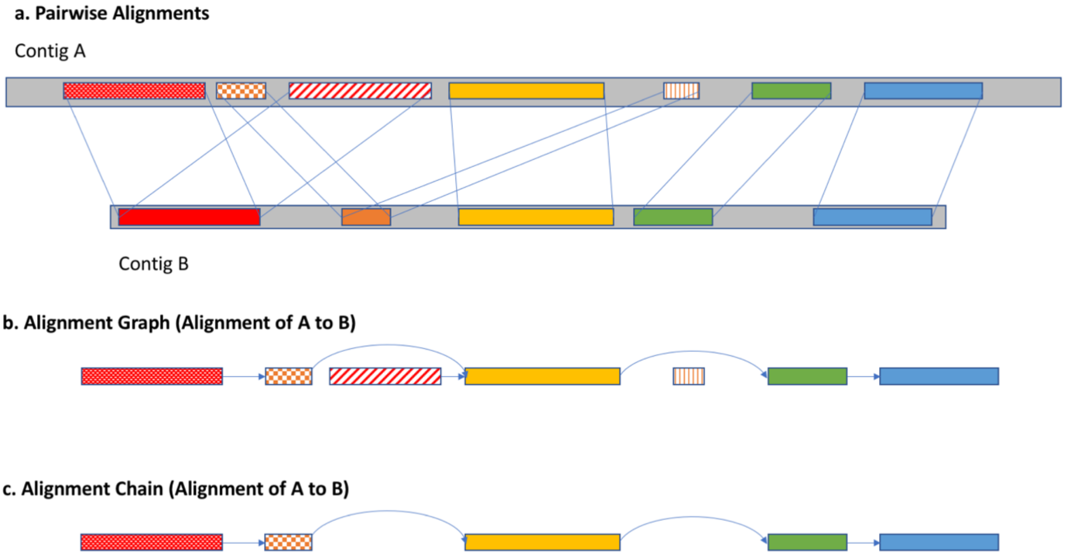
Supplementary Figure 21.25



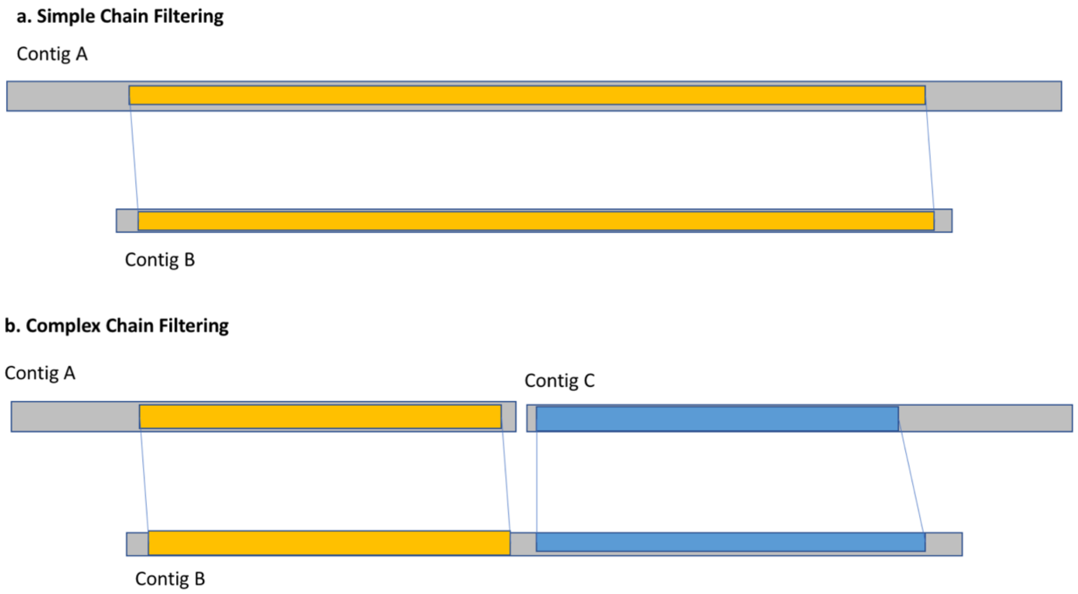
Supplementary Figure 22. Pseudo-haploid genome assembly. a, The original sample has two homologous chromosomes labeled orange and blue. **b**, In the de novo assembly, homologous regions containing higher rates of heterozygosity are split into distinct sequences (orange and blue), while regions with low rates or no heterozygous bases are collapsed to a single representative sequence (black). **c**, Our algorithm attempts to filter out redundant contigs from the other homologous chromosome, although the phasing of the differ contigs may be inconsistent (one orange and one blue contig remains). Figure derived from (Weisenfeld et al., 2017).

References

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757-767 (2017).



Supplementary Figure 23. Alignment chain construction. a, Pairwise alignments between all contigs are computed with nucmer. Here we show just the alignments between contigs A and B. **b,** An alignment graph is computed where each aligned region of A forms a node, with edges between nodes that are compatible on the same strand, in the same order, and no more than 20 kb between them. **c,** The final alignment chain is selected from the alignment graph as the maximal weight path in the alignment graph.



Supplementary Figure 24. Chain filtering. a, In simple cases, short contigs (contig A) are filtering out by their alignment chains to longer non-redundant contigs (contig B). **b,** In complex cases, a contig (contig B) is filtered out because the total span of the alignment chains to multiple non-redundant contigs (contigs A and C) span more than 93% of the bases.

Supplementary Table 1. Contig level assembly of CB5 genome

Items	Statistics
total assembled size (bp)	513048691
Number of contigs	1970
N90 (bp)	132736
N80 (bp)	197994
N70 (bp)	267596
N60 (bp)	340816
N50 (bp)	426696
Average length (bp)	260430
Maximum length (bp)	2187246

Supplementary Table 2. BUSCO analysis of Genome completeness

Description	Number	Percentage (%)
Complete BUSCOs (C)	1273	92.6
Complete and single-copy BUSCOs (S)	1202	87.4
Complete and duplicated BUSCOs (D)	71	5.2
Fragmented BUSCOs (F)	52	3.8
Missing BUSCOs (M)	50	3.6
Total BUSCO groups searched	1375	100

Supplementary Table 3. Assessment of genome assemblies based on RNA-seq assembled transcripts

	Number	Total Length (bp)	Accuracy (%)	Bases covered by assembly (%)	Sequences covered by assembly (%)	With >90% sequence in same chromosome		With >50% sequence in same chromosome	
						Number	Percent (%)	Number	Percent (%)
RNA assembled transcripts	28,162	35,608,402	99.92	99.88	100	280,64	99.65	28,161	99.99

Supplementary Table 4. Assessment of genome consistency based on Illumina reads

Items	**CB5 assembly**
Number of reads	153,521,279
Data size (Gb)	30.7
Mapped bases (Gb)	30.2
Mapping rate (%)	98.47
Genome Length (Mbp)	513
Mean Depth	61.54
Coverage Rate (%)	99.51
Regions with low coverage (< 5 reads)	5,222,770
Percentage with low coverage (< 5 reads)	1.01%
Number of homozygous variants	78,620
Percentage of homozygous variants	0.015%

Supplementary Table 5. Chromosome scale assembly of CB5 genome

ChrID	No. of contigs	Length	ChrID	No. of contigs	Length
Chr01	101	25202984	Chr13	61	17455539
Chr02	93	24587321	Chr14	92	17444028
Chr03	70	22012828	Chr15	57	17254157
Chr04	84	21990858	Chr16	68	17104571
Chr05	88	21937371	Chr17	52	16995643
Chr06	89	21313134	Chr18	67	16885232
Chr07	67	20232783	Chr19	57	16278134
Chr08	83	19980813	Chr20	54	15990161
Chr09	58	19858224	Chr21	64	14773377
Chr10	89	19604086	Chr22	61	14486819
Chr11	68	18936500	Chr23	65	13572809
Chr12	69	18615160	Chr24	55	12992659
			Chr25	54	10273841

Total number of contigs (bp)	1970
Total length of contigs (bp)	513048691
Total number of anchored contigs	1766
Total length of chromosomes (bp)	455779032
Number of unanchored contigs	204
Length of unanchored contigs	57443759
Anchor rate (%)	88.80

Supplementary Table 6. Comparison of pineapple CB5 and 'F153' annotation

Items	CB5	'F153'
Total gene number	29412	27024
Average gene length (bp)	6014	4894
Total Gene length (Mb)	177	132
% of genome	35	35
Average CDS length (bp)	187	212
Average exon number per gene	5.9	5.84
Average exon length (bp)	199	252
Average intron length (bp)	987	706
Number of miRNA	130	139

Supplementary Table 7. Statistics of TEs in CB5 genome

	Number	Length (Mb)	% of repeats	% of genome
Total repeat fraction	1,168,191	383.17	100.00	74.66
Class I: Retroelement	441,193	245.96	64.19	47.92
LTR Retrotransposon	252,868	184.68	48.20	35.98
Ty1/Copia	36,612	22.90	5.98	4.46
Ty3/Gypsy	76,647	87.48	22.83	17.05
Other	139,609	74.30	19.39	14.48
Non-LTR Retrotransposon	104,267	40.29	10.51	7.85
LINE	80,909	37.10	9.68	7.23
SINE	23,358	3.19	0.83	0.62
Unclassified retroelement	84,058	20.99	5.48	4.09
Class II: DNA transposon	366,222	113.52	29.63	22.12
TIR			0.00	0.00
CMC [DTC]	40,474	20.71	5.40	4.04
hAT	36,850	9.32	2.43	1.82
Mutator	24,304	8.10	2.11	1.58
Tc1/Mariner	0	0.00	0.00	0.00
PIF/Harbinger	6,449	2.16	0.56	0.42
Other	258,145	73.23	19.11	14.27
Helitron	10,127	2.27	0.59	0.44
Tandem Repeats	304,362	13.56	3.54	2.64
Unknown	37,136	10.13	2.64	1.97

Supplementary Table 8. The number of *CesA* gene in different species

	CesA1	CesA2	CesA3	CesA4	CesA5	CesA6	CesA7	CesA8	CesA9	CesA10	CesA11
Arabidopsis thaliana	1	1	1	1	1	1	1	1	1	1	
Oryza sativa	1	1	1	1	1	1	1	1	1		1
Apostasia shenzhenica	1		2		1		1	2	2		
Carica papaya	1		1	1		2	1	1	1		
Vitis vinifera	1		1	1		2	3	1	1		
‘F153’	1		1	1		1	1	1	1		1
CB5	1		1	1		1	1	1	1		1

Supplementary Table 9. Anthocyanin biosynthetic genes in ‘F153’ and CB5

Early biosynthetic		
Arabidopsis thaliana	CB5	‘F153’
Early biosynthetic genes		
AtCHS (AT5G13930)	CB5.v30111400	Aco016200.1
	CB5.v30128750	Aco007699.1
	CB5.v30128790	Aco008872.1
	CB5.v30194890	
	CB5.v30217630	
AtCHI (AT3G55120)	CB5.v30035150	Aco014232.1
	CB5.v30039890	
AtF3H (AT3G51240)	CB5.v30131050	Aco018609.1
	CB5.v30189820	
AtF3'H (AT5G07990)	CB5.v30155360	Aco003885.1
	CB5.v30301090	Aco019275.1
	CB5.v30158890	Aco017169.1
	CB5.v30245570	
AtFLS1 (AT5G08640)		
AtFLS2 (AT5G63580)		
AtFLS3 (AT5G63590)		
AtFLS4 (AT5G63595)		
AtFLS5 (AT5G63600)		
AtFLS6 (AT5G43935)		
Late biosynthetic genes		
Arabidopsis thaliana	CB5	‘F153’
AtDFR (AT5G42800)	CB5.v30035730	Aco006769.1
	CB5.v30079180	
	CB5.v30112570	
AtANS (AT4G22880)		
AtUGT79B1 (AT5G54060)	CB5.v30069980	Aco023087.1
	CB5.v30256190	Aco024663.1
	CB5.v30261160	Aco023089.1
	CB5.v30314600	Aco005407.1
AtUGT75C1 (AT4G14090)	CB5.v30153050	Aco000446.1
	CB5.v30165770	Aco018141.1
		Aco005246.1
AtUGT78D2 (AT5G17050)		Aco012126.1

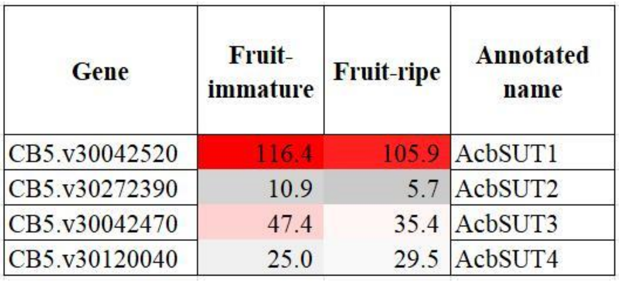
*Soure: *Arabidopsis thaliana* (Guo et al., 2014).

Guo, N., Cheng, F., Wu, J., Liu, B., Zheng, S.N., Liang, J.L., and Wang, X.W. Anthocyanin biosynthetic genes in *Brassica rapa*. *BMC Genomics* **15**, 426 (2014).

Supplementary Table 10. Numbers of sugar transporters and metabolism related genes in ‘F153’ and CB5

	Sweet	INV	UGPGP	SUS	SPS	SPP	SUT	STP	PMT	INT	Glc	SPF	VGT	TMT
‘F153’	18	4	5	9	4	1	4	18	7	3	3	5	2	3
CB5	20	7	7	11	5	1	4	23	10	3	3	8	2	4

Supplementary Table 11. The expression levels (Log$_2$ of FPKM values) of *SUTs* in different stages of CB5 fruits

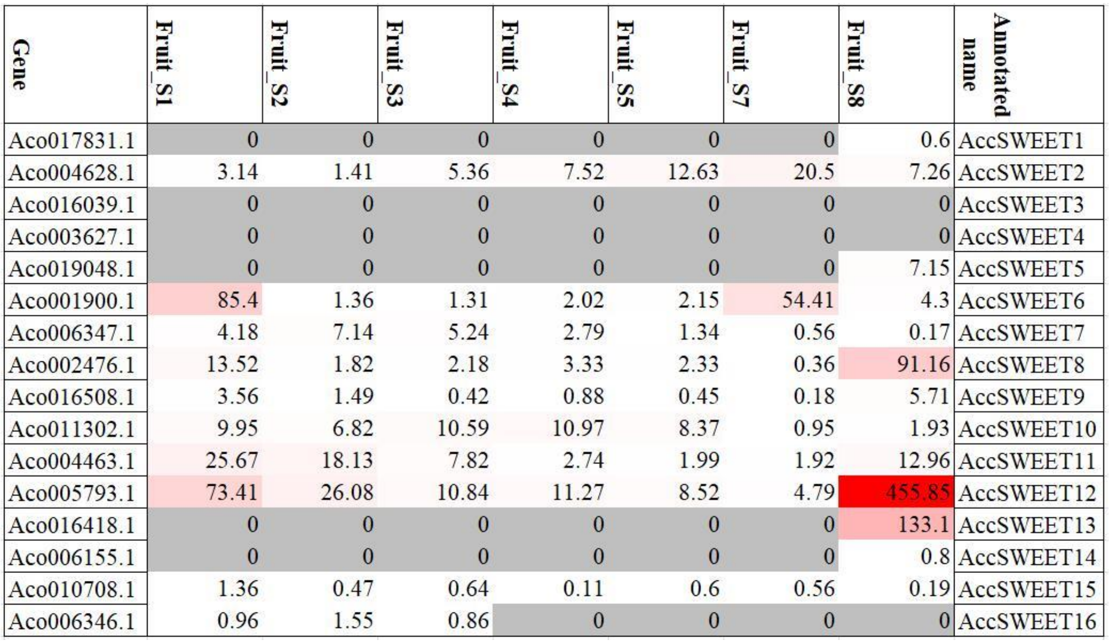


Gene	Fruit-immature	Fruit-ripe	Annotated name
CB5.v30042520	116.4	105.9	AcbSUT1
CB5.v30272390	10.9	5.7	AcbSUT2
CB5.v30042470	47.4	35.4	AcbSUT3
CB5.v30120040	25.0	29.5	AcbSUT4

Supplementary Table 12. The expression levels (Log$_2$ of FPKM values) of *SUTs* in different stages of MD2 fruits

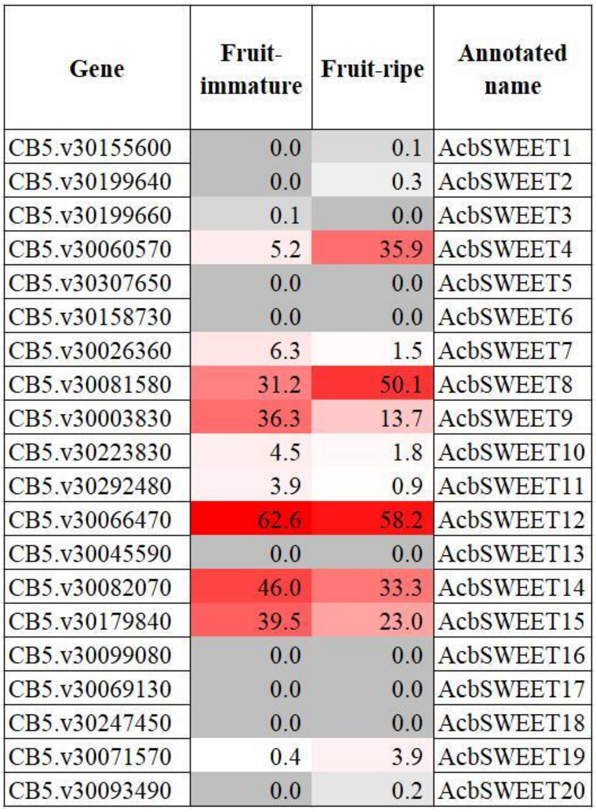
Gene	Fruit_S1	Fruit_S2	Fruit_S3	Fruit_S4	Fruit_S5	Fruit_S7	Fruit_S8	Annotated name
Aco004135.1	96.17	93.24	82.73	83.77	52.06	38.49	636.4	AccSUT1
Aco009281.1	7.25	6.13	6.89	2.86	2.82	1.16	10.82	AccSUT2
Aco004131.1	42.59	34.81	27.45	28.45	19.11	18.1	2720.58	AccSUT3
Aco000269.1	10.05	11.64	6.42	22.1	22.87	44.5	12.25	AccSUT4

Supplementary Table 13. The expression levels (Log_2 of FPKM values) of *SWEETs* in different stages of MD2 fruits



Gene	Fruit_S1	Fruit_S2	Fruit_S3	Fruit_S4	Fruit_S5	Fruit_S7	Fruit_S8	Annotated name
Aco017831.1	0	0	0	0	0	0	0.6	AccSWEET1
Aco004628.1	3.14	1.41	5.36	7.52	12.63	20.5	7.26	AccSWEET2
Aco016039.1	0	0	0	0	0	0	0	AccSWEET3
Aco003627.1	0	0	0	0	0	0	0	AccSWEET4
Aco019048.1	0	0	0	0	0	0	7.15	AccSWEET5
Aco001900.1	85.4	1.36	1.31	2.02	2.15	54.41	4.3	AccSWEET6
Aco006347.1	4.18	7.14	5.24	2.79	1.34	0.56	0.17	AccSWEET7
Aco002476.1	13.52	1.82	2.18	3.33	2.33	0.36	91.16	AccSWEET8
Aco016508.1	3.56	1.49	0.42	0.88	0.45	0.18	5.71	AccSWEET9
Aco011302.1	9.95	6.82	10.59	10.97	8.37	0.95	1.93	AccSWEET10
Aco004463.1	25.67	18.13	7.82	2.74	1.99	1.92	12.96	AccSWEET11
Aco005793.1	73.41	26.08	10.84	11.27	8.52	4.79	455.85	AccSWEET12
Aco016418.1	0	0	0	0	0	0	133.1	AccSWEET13
Aco006155.1	0	0	0	0	0	0	0.8	AccSWEET14
Aco010708.1	1.36	0.47	0.64	0.11	0.6	0.56	0.19	AccSWEET15
Aco006346.1	0.96	1.55	0.86	0	0	0	0	AccSWEET16

Supplementary Table 14. The expression levels (Log$_2$ of FPKM values) of *SWEETs* in different stages of CB5 fruits



Gene	Fruit-immature	Fruit-ripe	Annotated name
CB5.v30155600	0.0	0.1	AcbSWEET1
CB5.v30199640	0.0	0.3	AcbSWEET2
CB5.v30199660	0.1	0.0	AcbSWEET3
CB5.v30060570	5.2	35.9	AcbSWEET4
CB5.v30307650	0.0	0.0	AcbSWEET5
CB5.v30158730	0.0	0.0	AcbSWEET6
CB5.v30026360	6.3	1.5	AcbSWEET7
CB5.v30081580	31.2	50.1	AcbSWEET8
CB5.v30003830	36.3	13.7	AcbSWEET9
CB5.v30223830	4.5	1.8	AcbSWEET10
CB5.v30292480	3.9	0.9	AcbSWEET11
CB5.v30066470	62.6	58.2	AcbSWEET12
CB5.v30045590	0.0	0.0	AcbSWEET13
CB5.v30082070	46.0	33.3	AcbSWEET14
CB5.v30179840	39.5	23.0	AcbSWEET15
CB5.v30099080	0.0	0.0	AcbSWEET16
CB5.v30069130	0.0	0.0	AcbSWEET17
CB5.v30247450	0.0	0.0	AcbSWEET18
CB5.v30071570	0.4	3.9	AcbSWEET19
CB5.v30093490	0.0	0.2	AcbSWEET20

Supplementary Table 15. The number of bromelain genes in different species

Species	**Subfamily I**	**Subfamily II**	**Subfamily III**	**Subfamily IV**	**Subfamily V**	**Subfamily VI**	**Subfamily VII**	**Subfamily VIII**	**Subfamily IX**	**Total**
‘F153’	3	3	4	2	1	44	2	1	1	61
CB5	4	3	2	3	0	28	3	2	2	47
O. sativa	3	12	3	1	2	23	3	1	2	50
S. bicolor	8	10	3	1	1	19	3	0	2	47
P. trichocarpa	5	6	4	5	1	19	6	1	3	50
V. vinifera	3	2	2	2	1	8	5	1	1	25
C. papaya	3	2	10	3	1	10	3	1	1	34
A. thaliana	9	3	2	1	1	11	4	2	3	36
A. trichopoda	2	1	2	2	0	17	2	1	1	28

Supplementary Table 16. The expression levels (Log$_2$ of FPKM values) of bromelains in different stages of MD2 fruits

Gene	Fruit_S1	Fruit_S2	Fruit_S3	Fruit_S4	Fruit_S5	Fruit_S7	Fruit_S8	Subfamily	Annotated name
Aco011478.1	9.2	9.9	9.8	9.9	9.8	9.5	9.3	Subfamily I	AccRD21A
Aco018144.1	3.5	3.6	3.7	3.6	3.1	2.1	7.6	Subfamily I	AccRD21B
Aco021005.1	8.3	8.7	8.4	8.2	8.3	7.9	9.4	Subfamily I	AccRD21C
Aco003557.1	0.0	0.0	0.0	0.0	0.0	2.5	4.6	Subfamily Ⅱ	AccCEP1
Aco004517.1	0.0	0.2	0.3	0.5	0.1	0.1	1.6	Subfamily Ⅱ	AccCEP2
Aco017317.1	5.9	4.5	3.0	3.6	5.8	5.4	12.1	Subfamily Ⅱ	AccCEP3
Aco003574.1	1.3	0.4	0.7	0.6	0.0	0.3	2.7	Subfamily Ⅲ	AccXCP1
Aco013521.1	0.9	0.2	0.0	0.0	0.0	0.1	2.7	Subfamily Ⅲ	AccXCP2
Aco018520.1	1.5	1.2	1.9	1.9	1.7	1.1	0.2	Subfamily Ⅲ	AccXCP3
Aco026572.1	2.7	1.6	0.8	0.5	0.0	0.1	2.3	Subfamily Ⅲ	AccXCP4
Aco000385.1	3.7	3.2	3.2	1.7	2.4	3.3	1.9	Subfamily IV	AccXBCP3-1
Aco005269.1	5.4	5.8	5.9	6.1	6.5	6.0	7.2	Subfamily IV	AccXBCP3-2
Aco009735.1	2.7	2.2	1.4	0.7	0.7	0.5	0.0	Subfamily V	AccTHI1
Aco001368.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3	Subfamily VI	AccPAP1
Aco002011.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	Subfamily VI	AccPAP2
Aco004355.1	13.2	14.2	14.2	13.7	13.7	14.1	12.6	Subfamily VI	AccPAP3
Aco004356.1	14.1	14.3	14.4	13.6	13.6	14.3	12.9	Subfamily VI	AccPAP4
Aco004358.1	6.5	6.2	6.1	5.7	5.2	3.9	7.6	Subfamily VI	AccPAP5
Aco004359.1	5.0	4.9	4.7	4.3	3.1	1.4	7.5	Subfamily VI	AccPAP6
Aco004360.1	7.7	8.2	8.1	7.4	7.8	7.5	6.5	Subfamily VI	AccPAP7
Aco004830.1	0.8	0.0	0.0	0.0	0.0	0.0	7.3	Subfamily VI	AccPAP8
Aco006114.1	3.0	2.8	2.9	2.5	2.1	1.7	4.3	Subfamily VI	AccPAP9
Aco006117.1	0.0	0.0	0.2	0.1	0.2	1.1	0.0	Subfamily VI	AccPAP10
Aco007513.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP11
Aco008182.1	4.4	5.9	6.2	5.7	5.6	6.2	3.6	Subfamily VI	AccPAP12
Aco009864.1	3.7	3.2	3.1	2.3	1.8	0.6	2.7	Subfamily VI	AccPAP13
Aco009865.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP14
Aco010677.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP15
Aco012205.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP16
Aco012282.1	0.5	0.0	0.0	0.0	0.4	0.4	0.5	Subfamily VI	AccPAP17
Aco012283.1	0.5	0.0	0.3	0.0	0.4	0.1	0.0	Subfamily VI	AccPAP18
Aco012880.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP19
Aco012963.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP20
Aco014432.1	0.0	0.1	0.0	0.0	0.0	0.0	1.8	Subfamily VI	AccPAP21
Aco016478.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP22
Aco016479.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP23
Aco016485.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	Subfamily VI	AccPAP24
Aco017299.1	4.0	0.5	0.3	0.3	0.5	4.4	9.2	Subfamily VI	AccPAP25
Aco017558.1	0.0	0.0	0.0	0.0	0.0	0.0	8.5	Subfamily VI	AccPAP26
Aco017729.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP27
Aco018371.1	0.9	1.1	0.8	0.0	0.2	0.5	3.1	Subfamily VI	AccPAP28
Aco018375.1	0.0	0.0	0.7	0.0	0.0	0.3	1.6	Subfamily VI	AccPAP29
Aco018985.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP30
Aco018986.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP31
Aco018987.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP32
Aco024067.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP33
Aco024540.1	0.4	0.2	0.2	0.4	0.0	0.1	0.0	Subfamily VI	AccPAP34
Aco025188.1	0.8	0.0	0.0	0.0	0.0	0.0	1.7	Subfamily VI	AccPAP35
Aco025189.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	Subfamily VI	AccPAP36
Aco026830.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP37
Aco027656.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP38
Aco028703.1	0.5	0.0	0.3	0.0	0.4	0.1	0.0	Subfamily VI	AccPAP39
Aco028704.1	0.5	0.0	0.3	0.0	0.4	0.1	0.0	Subfamily VI	AccPAP40
Aco030565.1	1.4	0.6	0.4	0.3	0.0	0.0	1.1	Subfamily VI	AccPAP41
Aco031115.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP42
Aco031525.1	1.6	0.0	0.0	0.0	0.0	0.0	0.0	Subfamily VI	AccPAP43
Aco031765.1	0.0	0.0	0.8	0.0	0.9	0.0	0.9	Subfamily VI	AccPAP44
Aco012004.1	0.0	0.0	0.0	0.0	0.0	0.2	0.2	Subfamily VII	AccRD19A
Aco013510.1	11.2	11.8	11.5	11.2	11.2	12.2	11.5	Subfamily VII	AccRD19B
Aco019014.1	9.0	9.5	10.2	10.6	11.2	11.6	8.8	Subfamily VIII	AccAALP
Aco027767.1	8.2	8.1	8.2	8.6	9.2	9.4	7.5	Subfamily IX	AccCTB1

Supplementary Table 17. The expression levels (Log$_2$ of FPKM values) of bromelains in different stages of CB5 fruits

Gene	Fruit-immature	Fruit-ripe	Subfamily	Annotated name
CB5.v30047850	9.7	9.6	Subfamily I	AcbRD21A
CB5.v30153080	5.3	4.3	Subfamily I	AcbRD21B
CB5.v30236950	8.1	7.6	Subfamily I	AcbRD21C
CB5.v30238820	7.7	7.1	Subfamily I	AcbRD21D
CB5.v30049360	12.1	11.3	Subfamily II	AcbCEP1
CB5.v30102160	0.0	0.0	Subfamily II	AcbCEP2
CB5.v30279850	0.0	0.0	Subfamily II	AcbCEP3
CB5.v30027810	0.0	0.3	Subfamily III	AcbXCP1
CB5.v30229540	0.0	0.0	Subfamily III	AcbXCP2
CB5.v30030890	1.4	1.7	Subfamily IV	AcbXBCP3-1
CB5.v30182120	2.7	3.2	Subfamily IV	AcbXBCP3-2
CB5.v30305530	4.7	4.8	Subfamily IV	AcbXBCP3-3
CB5.v30010820	0.0	0.0	Subfamily VI	AcbPAP1
CB5.v30015930	0.0	0.0	Subfamily VI	AcbPAP2
CB5.v30022600	0.0	0.0	Subfamily VI	AcbPAP3
CB5.v30023470	0.0	0.0	Subfamily VI	AcbPAP4
CB5.v30023480	0.0	0.0	Subfamily VI	AcbPAP5
CB5.v30089440	0.0	0.0	Subfamily VI	AcbPAP6
CB5.v30103490	0.0	0.0	Subfamily VI	AcbPAP7
CB5.v30112960	0.0	0.0	Subfamily VI	AcbPAP8
CB5.v30125110	0.0	0.0	Subfamily VI	AcbPAP9
CB5.v30125120	3.3	3.5	Subfamily VI	AcbPAP10
CB5.v30125130	0.0	0.0	Subfamily VI	AcbPAP11
CB5.v30125920	0.6	0.2	Subfamily VI	AcbPAP12
CB5.v30127040	0.0	0.0	Subfamily VI	AcbPAP13
CB5.v30129600	0.0	0.0	Subfamily VI	AcbPAP14
CB5.v30137240	0.0	0.0	Subfamily VI	AcbPAP15
CB5.v30137260	0.0	0.0	Subfamily VI	AcbPAP16
CB5.v30149190	2.6	1.9	Subfamily VI	AcbPAP17
CB5.v30201610	0.0	0.0	Subfamily VI	AcbPAP18
CB5.v30217910	0.0	0.0	Subfamily VI	AcbPAP19
CB5.v30217920	0.0	0.0	Subfamily VI	AcbPAP20
CB5.v30225590	0.0	0.0	Subfamily VI	AcbPAP21
CB5.v30225600	0.0	0.0	Subfamily VI	AcbPAP22
CB5.v30229310	0.0	0.0	Subfamily VI	AcbPAP23
CB5.v30229320	0.0	0.0	Subfamily VI	AcbPAP24
CB5.v30237080	0.0	0.0	Subfamily VI	AcbPAP25
CB5.v30239060	0.0	0.0	Subfamily VI	AcbPAP26
CB5.v30250420	0.0	0.0	Subfamily VI	AcbPAP27
CB5.v30311080	0.0	0.0	Subfamily VI	AcbPAP28
CB5.v30011610	11.0	10.9	Subfamily VII	AcbRD19A
CB5.v30083280	0.0	0.0	Subfamily VII	AcbRD19B
CB5.v30205430	0.0	0.0	Subfamily VII	AcbRD19C
CB5.v30121320	7.4	7.0	Subfamily VIII	AcbALP2
CB5.v30241270	8.2	8.2	Subfamily VIII	AcbAALP
CB5.v30215110	8.1	8.3	Subfamily IX	AcbCTB1
CB5.v30228490	9.3	9.4	Subfamily IX	AcbCTB2

Supplementary Table 18. Summary of sequenced pineapple cultivars and wild *Ananas* accessions

Accession	Genus species	Common name	USDA accession	Original Classification	Sup Fig 10 color	Admixture analysis	Collection site
A01	*Ananas* species	DEW 1301	HANA173		mix 3 colors		Bolivia
Aa02	*A. comosus var. microstachys*	CB6	HANA65		green	*var.* microstachys	Brazil
Aa03	*A. comosus var. microstachys*	CB10	HANA67		green	*var.* microstachys	Brazil
Aa04	*A. comosus var. microstachys*	CB15	HANA69		green	*var.* microstachys	Paraguay
Aa05	*A. comosus var. microstachys*	CB19	HANA72		green	*var.* microstachys	Paraguay
Aa06	*A. comosus var. microstachys*	CB61	HANA83		mix 3 colors	Admixture	Brazil
Aa07	*A. comosus var. microstachys*	CB63	HANA84		green	*var.* microstachys	Brazil
Aa08	*A. comosus var. microstachys*	CB71	HANA88		green	*var.* microstachys	Brazil
Ab09	*A. comosus var. bracteatus*	F0-101	HANA20		mix 3 colors	Admixture	Brazil
Ab10	*A. comosus var. bracteatus*	CB5	HANA64		mix 4 colors	Admixture	Brazil
Ab11	*A. comosus var. bracteatus*	CB11	HANA68		mix 2 colors	Admixture	Brazil
Ab12	*A. comosus var. bracteatus*	CB17	HANA70		mix 2 colors	Admixture	Paraguay
Ab13	*A. comosus var. bracteatus*	CB20	HANA73		orange	*var.* bracteatus	Paraguay
Ab14	*A. comosus var. bracteatus*	CB21	HANA74		orange	*var.* bracteatus	Paraguay
Ab15	*A. comosus var. bracteatus*	CB23	HANA75		orange	*var.* bracteatus	Argentina
Ab16	*A. comosus var. bracteatus*	Albus	HANA82		mix 2 colors		Brazil
Ac17c	*A. comosus var. comosus*	41-411	HANA1	Hybrid	mix 3 colors	Admixture	Breeding program - Hawaii
Ac18c	*A. comosus var. comosus*	61-2223	HANA3	Cayenne	mix 4 colors	Admixture	Breeding program - Hawaii
Ac19s	*A. comosus var. comosus*	Spanish-Samoa	HANA13	S. Spanish	dark blue	Queen	American Samoa

Ac20q	*A. comosus var. comosus*	Mauritius	HANA18	Queen	dark blue	Queen	Taiwan
Ac21c	*A. comosus var. comosus*	Sarawak	HANA19	Cayenne	mix 4 colors	Admixture	Taiwan
Ac22p	*A. comosus var. comosus*	Abacaxi	HANA21	Abacaxi	mix 3 colors	Admixture	Brazil
Ac23s	*A. comosus var. comosus*	Wild-Kailua	HANA27	Cayenne	purple	S. Spanish	Hawaii
Ac24s	*A. comosus var. comosus*	Sugarloaf	HANA29	S. Spanish	purple	S. Spanish	Philippines
Ac25s	*A. comosus var. comosus*	Black Antigua	HANA31		purple	S. Spanish	Philippines
Ac26m	*A. comosus var. comosus*	Criolla	HANA37		yellow	Mordilona-related	Mexico
Ac27s	*A. comosus var. comosus*	Phu Qui	HANA40		purple	S. Spanish	Vietnam
Ac28	*A. comosus var. comosus*	Saigon Red	HANA42		purple	S. Spanish	Vietnam
Ac29	*A. comosus var. comosus*	Canterra	HANA47	S. Spanish	mix 4 colors	Admixture	Colombia
Ac30	*A. comosus var. comosus*	Bogota	HANA50	From Colombia	green	*var.* microstachys	Colombia
Ac31	*A. comosus var. comosus*	British-Samoa-P1	HANA53		mix 3 colors	Admixture	Samoa
Ac32s	*A. comosus var. comosus*	Spanish Guatemala	HANA60	Red Spanish	Yellow	Mordilona-related	Guatemala
Ac33c	*A. comosus var. comosus*	Cayenne M267 Dry sweet	HANA119	Cayenne	light blue	Cayenne	Hawaii
Ac34c	*A. comosus var. comosus*	Los Banos	HANA120		light blue	Cayenne	
Ac35	*A. comosus var. comosus*	Amarillo	HANA121	Cayenne	mix 4 colors	Admixture	Brazil
Ac36m	*A. comosus var. comosus*	Taboga	HANA124		Yellow	Mordilona-related	Panama
Ac37	*A. comosus var. comosus*	Smooth Anpi	HANA126		mix 4 colors	Admixture	Taiwan
Ac38c	*A. comosus var. comosus*	Cheese Pine	HANA132	Cayenne	mix 3 colors	Admixture	Guatemala
Ac39c	*A. comosus var. comosus*	Kew-1	HANA133	Cayenne	mix 3 colors	Admixture	Philippines
Ac40s	*A. comosus var. comosus*	Spanish Criolla Red	HANA136	S. Spanish	mix 4 colors	Admixture	Venezuela
Ac41c	*A. comosus var. comosus*	58-696	HANA156	Hybrid	mix 4 colors	Admixture	Breeding program -

							Hawaii
Ac42h	*A. comosus var. comosus*	63-759	HANA157	Hybrid	mix 2 colors	Admixture	Breeding program - Hawaii
Ac43h	*A. comosus var. comosus*	57-503	HANA158	Hybrid	light blue	Cayenne	Breeding program - Hawaii
Ac44c	*A. comosus var. comosus*	58-474	HANA161	Cayenne	light blue	Cayenne	Breeding program - Hawaii
Ac45h	*A. comosus var. comosus*	Champaka	HANA178	Hybrid	light blue	Cayenne	Breeding program - Hawaii
Ac46h	*A. comosus var. comosus*	Del monte gold		PRI Hybrid	mix 2 colors	Admixture	Breeding program - Hawaii
Ac47h	*A. comosus var. comosus*	Paull		PRI Hybrid	mix 2 colors	Admixture	Breeding program - Hawaii
Ac48q	*A. comosus var. comosus*	Comte de Paris	china	Queen	dark blue	Queen	China**
Ac49q	*A. comosus var. comosus*	Yellow Mauritius	china	Queen	dark blue	Queen	China
Ac50q	*A. comosus var. comosus*	Sarawak	china	Cayenne	dark blue	Queen	China
Ae51	*A. comosus var. erectifolius*	CB2	HANA63		mix 4 colors	Admixture	Brazil
Ae52	*A. comosus var. erectifolius*	CB9	HANA66		mix 4 colors	Admixture	Brazil
Ac53	*A. comosus var. bracteatus*	N04-8	HANA185		orange	*var.* bracteatus	Brazil
Pg54	*Pitcairnia gracilus*	N04-5	HANA182		brown	Pitcairnia sp.	South America
Pp55	*Pitcairnia punicea*	N04-6	HANA183		brown	Pitcairnia sp	South America
Ac56m	*A. comosus var. comosus*	Monte Lirio	HANA34	Mordilona	yellow	Mordilona-related	Guatemala
Ac57q	*A. comosus var. comosus*	MacGregor	HANA25	Queen	dark blue	Queen	Australia
Ac58s	*A. comosus var. comosus*	Ruby	HANA15	S. Spanish	purple	S. Spanish	Singapore
Ac59p	*A. comosus var. comosus*	Abacaxi vermelho	HANA147	Pernambuco	mix 4 colors	Admixture	Brazil
Ac60s	*A. comosus var. comosus*	Red Spanish	HANA123	Red Spanish	light blue	Cayenne	Panama

Ac61q	*A. comosus var. comosus*	Vietnam queen 3	china	Queen	dark blue	Queen	China
Ac62q	*A. comosus var. comosus*	Pulae	china	S. Spanish	dark blue	Queen	China
Ac63s	*A. comosus var. comosus*	Spanish red spine	china	S. Spanish	purple	S. Spanish	China
Ac64s	*A. comosus var. comosus*	Ornamental queen	china	S. Spanish	purple	S. Spanish	China
Ac65m	*A. comosus var. comosus*	Cambray	HANA32	Mordilona	yellow	Mordilona-related	Philippines
Ac66s	*A. comosus var. comosus*	Ananas Merah	HANA131	S. Spanish	purple	S. Spanish	Indonesia
Ac67s	*A. comosus var. comosus*	Congo	HANA12	S. Spanish	purple	S. Spanish	Congo, Africa
Ac68q	*A. comosus var. comosus*	Nenas Arnis	china	Queen	dark blue	Queen	China
Ac69s	*A. comosus var. comosus*	Sylhet jaldubi	HANA30	S. Spanish	purple	S. Spanish	Philppines
Ac70q	*A. comosus var. comosus*	Vietnam queen 2	china	Queen	dark blue	Queen	China
Ac71q	*A. comosus var. comosus*	Vietnam queen 1	china	Queen	dark blue	Queen	China
Ac72q	*A. comosus var. comosus*	James Queen	china	Tetraploid Queen	dark blue	Queen	China
Ac73p	*A. comosus var. comosus*	Pérola	china	Pernambuco	mix 3 colors	Admixture	China
Ac74q	*A. comosus var. comosus*	India queen	china	Queen	dark blue	Queen	China
Ac75m	*A. comosus var. comosus*	Papuri vaupes Colombia	HANA52	Mordilona	mix 3 colors	Admixture	Colombia
Ac76s	*A. comosus var. comosus*	Uhi	HANA122	S. Spanish	purple	S. Spanish	Taiwan
Ac77m	*A. comosus var. comosus*	Rondon	HANA81	Mordilona	mix 2 colors	Admixture	Brazil
Ac79s	*A. comosus var. comosus*	MO	HANA43	S. Spanish	purple	S. Spanish	Vietnam
Ac80c	*A. comosus var. comosus*	Cayenne hilo	HANA10	Cayenne	light blue	Cayenne	Hawaii
Ac81c	*A. comosus var. comosus*	Cayenne azores	HANA139	Cayenne	light blue	Cayenne	Azores Islands, Spain
Ac82	*A. comosus var. comosus*	White Jade	HANA155		mix 2 colors	Admixture	Hawaii
Ac83s	*A. comosus var. comosus*	Philippine red	HANA26		purple	S. Spanish	Philippines

Ac84s	*A. comosus var. comosus*	Cayenne lanai	HANA8	Cayenne	purple	S. Spanish	Hawaii
Ac85s	*A. comosus var. comosus*	Philippine green	HANA129	S. Spanish	purple	S. Spanish	Philippines
Ac86m	*A. comosus var. comosus*	Morada	HANA135		mix 2 colors	Admixture	Venezuela
Ac87s	*A. comosus var. comosus*	Amalsad	HANA35	S. Spanish	orange	S. Spanish	Brazil
Aa88	*A. comosus var. microstachys*	Wild brazil	HANA38		green	*var.* microstachys	India
Ac89s	*A. comosus var. comosus*	Pho lang tuang	HANA41	S. Spanish	purple	S. Spanish	Vietnam

* Colors correlate with groupings from admixture analysis in Fig. 3.
** The collection in China is from arboretums not the wild.

Supplementary Table 19. Summary resequencing statistics for the 89 resequenced genomes

Accession	Sequence coverage	Sequencing mode	No. of reads (1)	No. of paired end reads
A01	3.6	paired end 150nt	6,283,757	12,567,514
Aa02	29.2	paired end 150nt	51,144,946	102,289,892
Aa03	10.9	paired end 150nt	18,991,680	37,983,360
Aa04	11.1	paired end 150nt	19,398,196	38,796,392
Aa05	14.2	paired end 150nt	24,855,312	49,710,624
Aa06	17.1	paired end 150nt	29,962,451	59,924,902
Aa07	14.1	paired end 150nt	24,596,559	49,193,118
Aa08	13.2	paired end 150nt	23,109,208	46,218,416
Ab09	8.9	paired end 150nt	15,531,310	31,062,620
Ab10	50.6	paired end 150nt	88,562,449	177,124,898
Ab11	20.1	paired end 150nt	35,224,705	70,449,410
Ab12	15.1	paired end 150nt	26,457,330	52,914,660
Ab13	22.7	paired end 150nt	39,759,495	79,518,990
Ab14	14.9	paired end 150nt	26,046,996	52,093,992
Ab15	13.3	paired end 150nt	23,250,828	46,501,656
Ab16	10.2	paired end 150nt	17,811,692	35,623,384
Ac17c	34.7	paired end 150nt	60,678,334	121,356,668
Ac18c	23.3	paired end 150nt	40,805,220	81,610,440
Ac19s	17	paired end 150nt	29,663,155	59,326,310
Ac20q	2.8	paired end 150nt	4,891,288	9,782,576
Ac21c	11.2	paired end 150nt	19,663,867	39,327,734
Ac22a	42.6	paired end 150nt	74,582,158	149,164,316
Ac23c	14.1	paired end 150nt	24,710,571	49,421,142
Ac24a	7.6	paired end 150nt	13,348,850	26,697,700
Ac25	7.8	paired end 150nt	13,594,157	27,188,314
Ac26	19.3	paired end 150nt	33,812,624	67,625,248
Ac27	14.8	paired end 150nt	25,906,288	51,812,576
Ac28	21.6	paired end 150nt	37,720,119	75,440,238
Ac29	21.0	paired end 150nt	36,748,704	73,497,408
Ac30	21.7	paired end 150nt	38,024,676	76,049,352
Ac31	13.6	paired end 150nt	23,775,777	47,551,554
Ac32s	18.6	paired end 150nt	32,518,762	65,037,524
Ac33c	7.7	paired end 150nt	13,457,821	26,915,642
Ac34	9.3	paired end 150nt	16,250,800	32,501,600
Ac35	10.5	paired end 150nt	18,460,855	36,921,710
Ac36	10.9	paired end 150nt	19,060,010	38,120,020
Ac37	8.1	paired end 150nt	14,209,224	28,418,448
Ac38c	10.4	paired end 150nt	18,152,904	36,305,808
Ac39c	8.5	paired end 150nt	14,908,500	29,817,000
Ac40s	10.8	paired end 150nt	18,864,220	37,728,440
Ac41c	6.9	paired end 150nt	12,062,392	24,124,784
Ac42c	4.1	paired end 150nt	7,120,229	14,240,458
Ac43c	5.1	paired end 150nt	8,976,945	17,953,890
Ac44c	6.8	paired end 150nt	11,984,521	23,969,042

Ac45c	5.6	paired end 150nt	9,866,699	19,733,398
Ac46c	9	paired end 150nt	15,764,135	31,528,270
Ac47c	10.8	paired end 150nt	18,941,881	37,883,762
Ac48	25.9	paired end 150nt	45,331,167	90,662,334
Ac49	33.4	paired end 150nt	58,377,159	116,754,318
Ac50	23.4	paired end 150nt	40,987,591	81,975,182
Ae51	30.7	paired end 150nt	53,759,917	107,519,834
Ae52	10.4	paired end 150nt	18,224,080	36,448,160
Cs53	26.6	paired end 150nt	46,496,849	92,993,698
Pg54	21	paired end 150nt	36,681,492	73,362,984
Pp55	40.5	paired end 150nt	70,805,015	141,610,030
Ac56m	38	paired end 250nt	39,920,551	79,841,102
Ac57q	38.4	paired end 250nt	40,292,387	80,584,774
Ac58s	31.8	paired end 250nt	33,433,855	66,867,710
Ac59p	38.5	paired end 250nt	40,455,395	80,910,790
Ac60s	38.4	paired end 250nt	40,334,959	80,669,918
Ac61q	32.2	paired end 250nt	33,841,329	67,682,658
Ac62s	33.5	paired end 250nt	35,127,307	70,254,614
Ac63s	35.6	paired end 250nt	37,404,491	74,808,982
Ac64s	32.5	paired end 250nt	34,149,048	68,298,096
Ac65m	31.2	paired end 250nt	32,711,977	65,423,954
Ac66s	38.9	paired end 250nt	40,886,838	81,773,676
Ac67s	27.1	paired end 250nt	28,476,959	56,953,918
Ac68s	23.7	paired end 250nt	24,905,409	49,810,818
Ac69s	32.8	paired end 250nt	34,474,549	68,949,098
Ac70q	26.9	paired end 250nt	28,207,332	56,414,664
Ac71q	28.2	paired end 250nt	29,661,646	59,323,292
Ac72q	29.7	paired end 250nt	31,195,261	62,390,522
Ac73p	30.4	paired end 250nt	31,968,735	63,937,470
Ac74p	33.4	paired end 250nt	35,019,347	70,038,694
Ac75m	8.9	paired end 250nt	9,306,564	18,613,128
Ac76s	7.2	paired end 250nt	7,599,296	15,198,592
Ac77m	8.1	paired end 250nt	8,514,748	17,029,496
Ac79s	8.1	paired end 250nt	8,497,456	16,994,912
Ac80	8.3	paired end, 250nt	8,718,417	17,436,834
Ac81	13.2	paired end, 250nt	13,848,418	27,696,836
Ac82	11.4	paired end, 250nt	11,937,427	23,874,854
Ac83	11.6	paired end, 250nt	13,278,151	26,556,302
Ac84	10	paired end, 250nt	10,460,545	20,921,090
Ac85s	8.5	paired end, 250nt	8,969,158	17,938,316
Ac86m	11.9	paired end, 250nt	12,466,234	24,932,468
Ac87s	8.4	paired end, 250nt	8,822,046	17,644,092
Aa88	10.8	paired end, 250nt	11,295,629	22,591,258
Ac89s	7.1	paired end, 250nt	7,493,569	14,987,138

Supplementary Table 20. The list of predicted effects of annotated SNPs

Type of predicted effect	Number of SNPs	Affected Gene No.
Large-effect SNPs		
SNPs that introduce stop codons	7,084	4,673
SNPs that disrupt stop codons	725	687
SNPs that disrupt start codons	750	688
SNPs that affect splice sites	4,252	3,188
Total large-effect SNPs	12,806	7,284
Synonymous SNPs within coding regions	689,019	21,190
Non-Synonymous SNPs within coding regions	589,484	21,453

Supplementary Table 21. List of number of non-synonymous (dN) and synonymous (dS) SNPs and dN/dS for each cultivar group

	Non-synonymous	Synonymous	dN/dS
Ananas comosus var. comosus	214,849	226,715	0.947661
Ananas comosus var. microstachys	182,209	192,779	0.94517
Ananas comosus var. erectifolius	86,064	95,799	0.898381
Ananas comosus var. bracteatus	185,642	207,110	0.896345
Cayenne	137,144	150,456	0.911522
Singapore Spanish	139,045	150,990	0.920889
Queen	82,444	88,329	0.933374

Supplementary Table 22. The number of transposable element junctions, belonging to four families in the 89 pineapple accessions that are shared with the reference pineapple genome ‘F153’

	Number				Percent			
Accessions	*Mutator*	PIF/Harbinger	hAT	CACTA	% Mutator	% PIF/Harbinger	% hAT	% CACTA
F153	46613	23634	18831	4091	100	100	100	100
Ac50	44534	22951	18290	3966	96	97	97	97
Ac17c	42763	21743	17667	3551	92	92	94	87
Ac22a	40808	20843	16824	3245	88	88	89	79
Ac19s	40722	20789	16839	3337	87	88	89	82
Ac45c	39810	21036	16863	3618	85	89	90	88
Ac49	38906	20042	16202	3211	83	85	86	78
Ac60s	38896	21713	16895	3784	83	92	90	92
Ac44c	38886	20538	15967	3332	83	87	85	81
Ac21c	38853	19982	16157	3113	83	85	86	76
Ac40s	38362	19758	15937	3072	82	84	85	75
Ac43c	37408	19899	15634	3173	80	84	83	78
Ac41c	36893	19488	15366	3072	79	82	82	75
Ac35	35777	19582	15275	3254	77	83	81	80
Ac62s	35645	19039	14894	3087	76	81	79	75
Ac61q	35554	19020	14836	3103	76	80	79	76
Ac59p	35519	19909	15657	3293	76	84	83	80
Ac64s	35433	18726	14675	3018	76	79	78	74
Ac82	35285	19119	14791	3178	76	81	79	78
Ac68s	35216	18774	14585	3052	76	79	77	75
Ac34	34772	18811	14747	2972	75	80	78	73
Ab10	34766	19425	15391	3132	75	82	82	77
Ac70q	34744	18544	14459	3019	75	78	77	74
Ac42c	34501	18073	14171	2889	74	76	75	71
Ac56m	34351	18276	14430	2871	74	77	77	70
Ac24a	34266	18391	14360	2874	74	78	76	70
Ac63s	34225	18356	14409	2971	73	78	77	73
Ac48	34128	18750	15052	2934	73	79	80	72
Ac27	34113	18571	14707	2900	73	79	78	71
Ac29	33979	18681	14735	2923	73	79	78	71
Ac67s	33932	18161	14249	2940	73	77	76	72
Ac57q	33152	18752	14779	3053	71	79	78	75
Ac65m	32863	17553	13733	2788	71	74	73	68

Ac71q	32498	17752	13817	2892	70	75	73	71
Ac18c	32180	18405	14807	3017	69	78	79	74
Ac25	29687	16887	13141	2630	64	71	70	64
Ab13	29521	16326	13296	2346	63	69	71	57
Ac38c	29418	17027	13227	2625	63	72	70	64
Ac72q	29357	16305	12514	2635	63	69	66	64
Ac23c	29003	16843	13592	2642	62	71	72	65
Ac58s	28507	16923	13293	2762	61	72	71	68
Ac69s	28324	16159	12714	2647	61	68	68	65
Ac28	28175	16810	13496	2742	60	71	72	67
Ab09	27964	16130	13169	2488	60	68	70	61
Ab12	27600	15447	12297	2255	59	65	65	55
Ac26	26959	16300	13306	2354	58	69	71	58
Ac33c	26645	16560	13043	2970	57	70	69	73
Ac73p	26473	15351	11880	2485	57	65	63	61
Aa02	26020	14422	11697	1736	56	61	62	42
Ac32s	25799	15884	12885	2361	55	67	68	58
Ac81	25682	15087	11868	2599	55	64	63	64
Cs53	24559	15365	12716	2165	53	65	68	53
Aa08	24321	14243	11638	1694	52	60	62	41
Ac30	24317	13806	11121	1629	52	58	59	40
A01	23267	12639	10260	1532	50	53	54	37
Ab14	22848	14286	11422	2084	49	60	61	51
Ae51	22762	14916	12421	1988	49	63	66	49
Ac66s	22181	15269	12317	2412	48	65	65	59
Ae52	21760	13551	11071	1830	47	57	59	45
Ab15	21565	12989	10811	1861	46	55	57	45
Ac74p	21208	13337	10155	2187	45	56	54	53
Ac31	20987	13413	11553	1966	45	57	61	48
Ac80	20698	14400	11526	2432	44	61	61	59
Ac36	20445	13159	11163	1944	44	56	59	48
Aa05	20268	12308	9995	1484	43	52	53	36
Aa07	20073	12254	9892	1478	43	52	53	36
Aa06	19588	12260	10209	1491	42	52	54	36
Ab11	19348	12230	10500	1681	42	52	56	41
Ac86m	19114	12643	9707	1881	41	53	52	46
Ac76s	18929	11824	9050	1894	41	50	48	46
Ac39c	18769	12578	11034	1851	40	53	59	45
Ab16	18666	11392	9632	1525	40	48	51	37
Ac37	18614	12730	10587	1924	40	54	56	47
Aa03	17739	11197	9531	1324	38	47	51	32
Ac79s	17089	10405	7987	1667	37	44	42	41
Ac83	16878	10346	8593	1638	36	44	46	40
Ac77m	16629	10255	7815	1660	36	43	42	41
Ac85s	16549	11395	8702	1800	36	48	46	44
Ac47c	15927	11198	9866	1561	34	47	52	38

Ac75m	15922	9800	7557	1464	34	41	40	36
Ac87s	15209	9786	7703	1556	33	41	41	38
Aa04	14300	9501	8242	1108	31	40	44	27
Ac84	13093	9235	7974	1520	28	39	42	37
Ac89s	12494	8058	6331	1277	27	34	34	31
Ac46c	10648	8975	8455	1152	23	38	45	28
Aa88	8733	6028	4956	694	19	26	26	17
Ac20q	6167	4486	4301	572	13	19	23	14
Pg54	812	491	415	68	2	2	2	2
Pp55	752	607	503	34	2	3	3	1

Supplementary Table 23. **Summary of putative swept regions during pineapple domestication**

Linkage Group	Sweep start pos.	Sweep end pos.	Sweep size (bp)
LG03	575001	800000	225000
LG04	1225001	1750000	525000
LG04	9600001	9975000	375000
LG06	1450001	1700000	250000
LG06	8750001	8975000	225000
LG06	11775001	12200000	425000
LG07	975001	1375000	400000
LG07	2650001	3200000	550000
LG07	10825001	11600000	775000
LG07	12875001	13300000	425000
LG09	675001	1275000	600000
LG10	2325001	2625000	300000
LG10	6850001	7100000	250000
LG10	11625001	12825000	1200000
LG12	1475001	1900000	425000
LG15	4750001	5525000	775000
LG17	1725001	2100000	375000
LG17	5200001	5525000	325000
LG17	9550001	9700000	150000
LG17	9800001	10550000	750000
LG18	8000001	9150000	1150000
LG21	6525001	6875000	350000
LG22	1950001	2400000	450000
LG23	6900001	7125000	225000
LG24	825001	1250000	425000

Supplementary Table 24. *S-RNase* and *SFL/SFB* homologs in pineapple 'F153' and their transcript levels in floral organs

		FPKM value	
SI candidate gene	**Gene description**	**mature androecium**	**mature gynoecium**
Aco000868	F-box family protein	707.0	121.4
Aco001100	Ribonuclease T2 family protein	1236.6	215.7
Aco001170	Kelch repeat-containing protein	15.2	10.7
Aco003507	F-box family protein	89.1	31.0
Aco003647	F-box and associated interaction domains-containing protein	222.4	407.4
Aco004110	Kelch-like protein	517.9	428.3
Aco004148	ribonuclease 1	28.4	58.5
Aco004758	ribonuclease 2	2191.7	1628.0
Aco005545	F-box family protein	267.2	140.7
Aco007352	F-box family protein	0	0
Aco007541	F-box family protein	171.4	42.9
Aco010730	F-box/kelch-repeat protein	698.9	23.7
Aco010854	Sec14p-like phosphatidylinositol transfer family protein	4020.7	3335.7
Aco011265	F-box protein	730.3	28.9
Aco012216	lipid-binding serum glycoprotein family protein	320.3	135.4
Aco013003	F-box/kelch-repeat protein	632.6	488.5
Aco013971	F-box protein	10.9	10.9
Aco015095	F-box family protein	1650.0	2282.5
Aco017266	F-box protein	257.9	286.1
Aco018964	F-box family protein	372.8	445.6
Aco021447	F-box family protein	135.9	623.3
Aco021972	F-box family protein	0.4	0
Aco024998	F-box family protein	1.1	0.3
Aco027913	hypothetical protein	0	0
Aco031303	F-box family protein	89.2	59.1

Supplementary Table 25. ***S-RNase*** **and** ***SFL/SFB*** **homologs in CB5**

SI candidate gene	Gene description
CB5.v30007890	F-box protein
CB5.v30007910	F-box protein
CB5.v30013780	F-box kelch-repeat protein
CB5.v30014510	Ribonuclease
CB5.v30016070	F-box like protein
CB5.v30016460	F-box protein
CB5.v30042270	F-box kelch-repeat protein
CB5.v30042660	Ribonuclease
CB5.v30048360	F-box kelch-repeat protein
CB5.v30076510	F-box protein
CB5.v30080150	F-box associated domain protein
CB5.v30087920	F-box kelch-repeat protein
CB5.v30090030	F-box associated domain protein
CB5.v30098650	F-box protein
CB5.v30101650	F-box and associated interaction domains-containing protein
CB5.v30138150	F-box family protein
CB5.v30144350	F-box kelch-repeat protein
CB5.v30150070	F-box family protein
CB5.v30163520	F-box only protein
CB5.v30165790	Ribonuclease
CB5.v30174180	CRAL/TRIO, N-terminal domain
CB5.v30185550	F-box kelch-repeat protein
CB5.v30187610	F-box protein
CB5.v30191740	F-box and associated interaction domains-containing protein
CB5.v30200880	F-box kelch-repeat protein
CB5.v30231780	F-box and associated interaction domains-containing protein
CB5.v30255300	F-box and associated interaction domains-containing protein
CB5.v30282170	F-box protein
CB5.v30286630	F-box protein
CB5.v30288000	F-box kelch-repeat protein
CB5.v30294170	F-box kelch-repeat protein
CB5.v30304450	CRAL/TRIO, N-terminal domain
CB5.v30304460	SEC14 cytosolic factor

Supplementary Table 26. The list of RNA sequencing samples

RNA sequencing samples		Biological Replicates
Pineapple varieties	Tissues	
CB5	Mature leaf and flower	3
CB5	One-month old fruit and two-month old fruit	5
F153	Mature leaf, flower and root	1
F153	Mature androecium and gynoecium	3
MD2	8 stages of fruits (S1-S8)	1

Supplementary Table 27. Statistics of Hi-C mapping

	Statistics of mapping
Clean Paired-end Reads	146556124
Unmapped Paired-end Reads	28723897
Unmapped Paired-end Reads Rate (%)	19.599
Paired-end Reads with Singleton	81807952
Paired-end Reads with Singleton Rate(%)	55.82
Multi Mapped Paired-end Reads	11832681
Multi Mapped Ratio (%)	8.074
Unique Mapped Paired-end Reads	24191594
Unique Mapped Ratio (%)	16.507
	Statistics of valid reads
Unique Mapped Paired-end Reads	24191594
Dangling End Paired-end Reads	1658631
Dangling End Rate (%)	6.856
Self Circle Paired-end Reads	5426632
Self Circle Rate (%)	22.432
Dumped Paired-end Reads	7646699
Dumped Rate (%)	31.609
Interaction Paired-end Reads	9087439
Interaction Rate (%)	37.564
Lib Valid Paired-end Reads	8016862
Lib Valid Rate (%)	88.219
Lib Dup (%)	11.781

Supplementary Table 28. BUSCO analysis of annotation completeness

Description	Number	Percentage (%)
Complete BUSCOs (C)	1165	84.7
Complete and single-copy BUSCOs (S)	982	71.4
Complete and duplicated BUSCOs (D)	183	13.3
Fragmented BUSCOs (F)	111	8.1
Missing BUSCOs (M)	99	7.2
Total BUSCO groups searched	1375	100