Fig. S1 | 'EA78' genome survey and sequencing. (A) Genome size and ploidy estimation for 'EA78'. The left panel displays genome size estimates based on k-mers, while the middle panel shows estimates from flow cytometry, using *Oryza sativa* cv. 'Japonica' as internal control. The right panel presents polyploid estimation. **(B)** The distribution of read length and base quality generated by different long-reads sequencing platforms. **(C)** Scatter plots showing parent-specific k-mers identified in the sequenced long reads, while the size of each blob representing the size of the read. Reads are plotted based on the number of identified paternal-specific kmers (x-axis) and maternal-specific k-mers (y-axis).

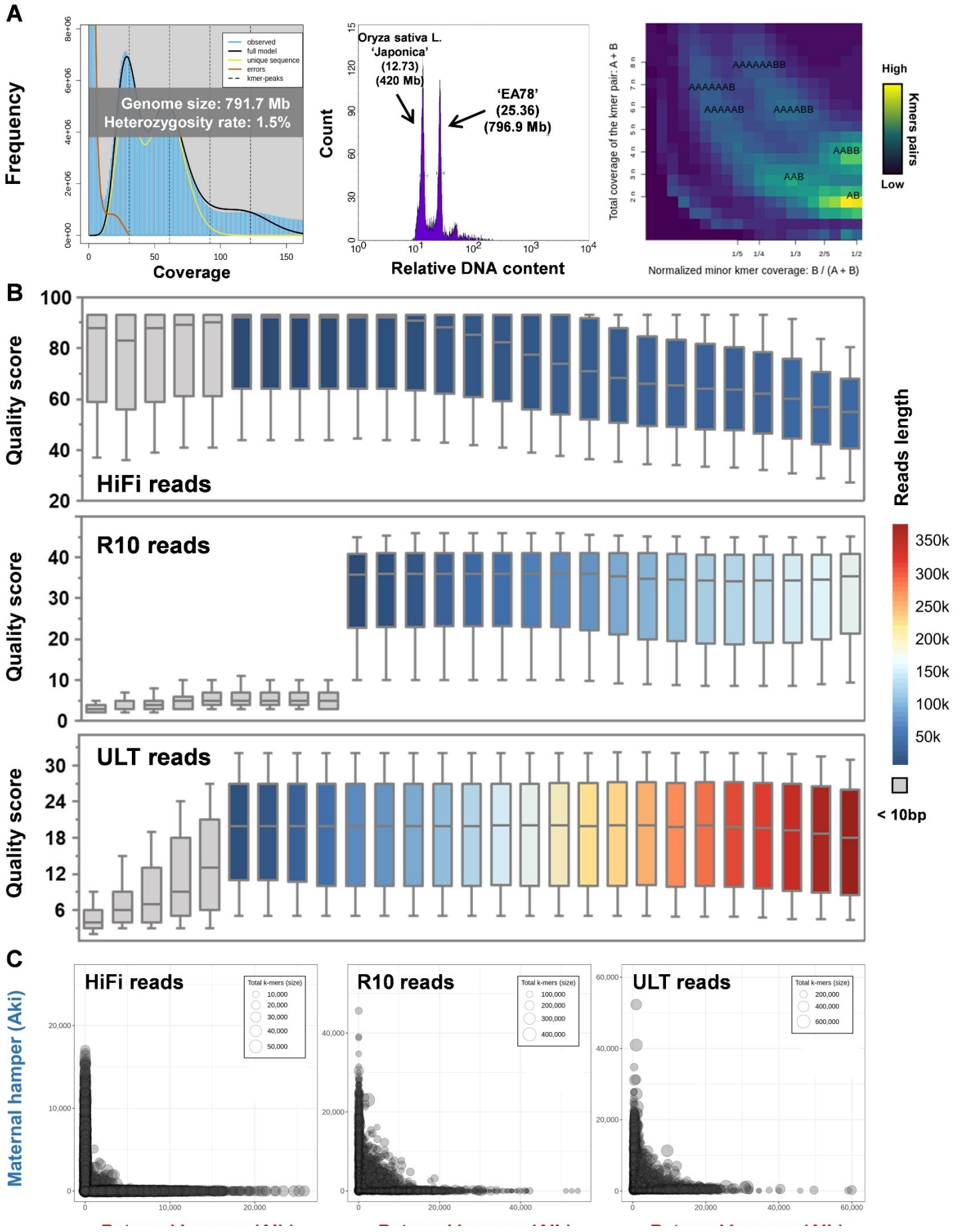


Fig. S2 | Evaluation of different assembly strategies for phasing the octoploid 'EA78' genome. Note: "HiFi only (primary)" represents an assembly strategy that relies solely on HiFi reads, ensuring the continuity of the primary assembly. "HiFi only (dual)" represents an assembly strategy that uses HiFi reads, while maintaining a certain level of continuity for both haplotypes. "HiFi + Hi-C" denotes a phased assembly strategy that integrates both HiFi reads and Hi-C reads. "HiFi + Trio" indicates a strategy that employs parental Illumina reads to differentiate paternal and maternal HiFi reads for phased assembly. Contig N90 metrics and hamming error rates are shown for each assembly. Red blobs represent contigs of the paternal haplotype (haplotype1), while blue blobs indicate the maternal haplotype (haplotype2). The size of each blob corresponds to the size of the contig, plotted according to the number of paternal-specific k-mers (x-axis) and maternal specific k-mers (y-axis) identified.

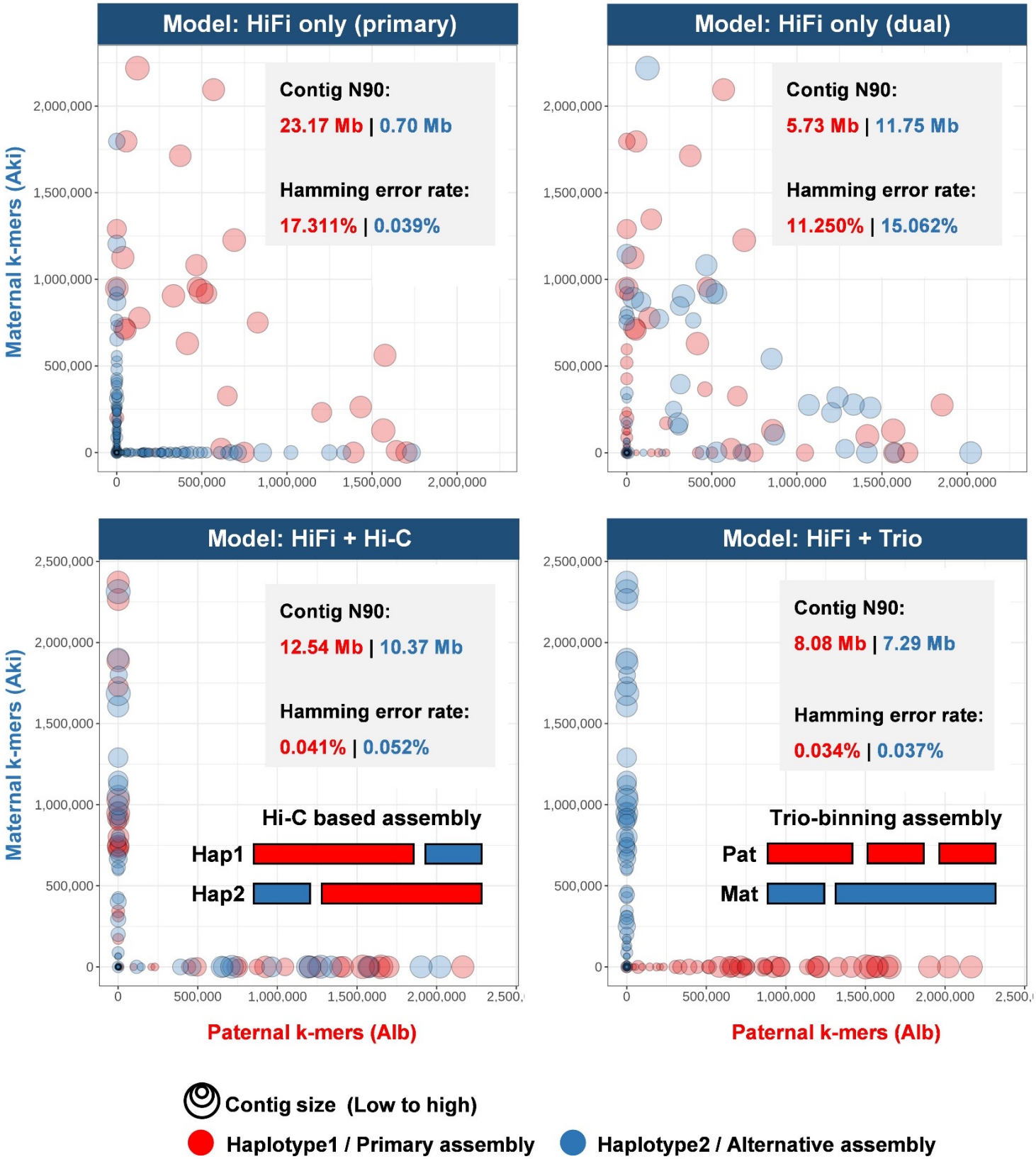


Fig. S3 | Evaluation of the effects of integrating Nanopore long reads on 'EA78' phased assemblies. (A) Comparison of genome assembly quality using Nanopore R10 reads (ONT R10) and Nanopore ultra-long reads (ONT UL) to assist phased assemblies. Note: Red blobs denote contigs of the haplotype1/paternal haplotype, while blue blobs indicate the haplotype2/maternal haplotype. The size of each blob corresponds to the size of the contig, plotted according to the number of paternal-specific k-mers (x-axis) and maternal-specific k-mers (y-axis) identified. **(B)** Assembly graphs generated from the combination of HiFi + Hi-C + ONT ULT and HiFi + Trio + ONT ULT.

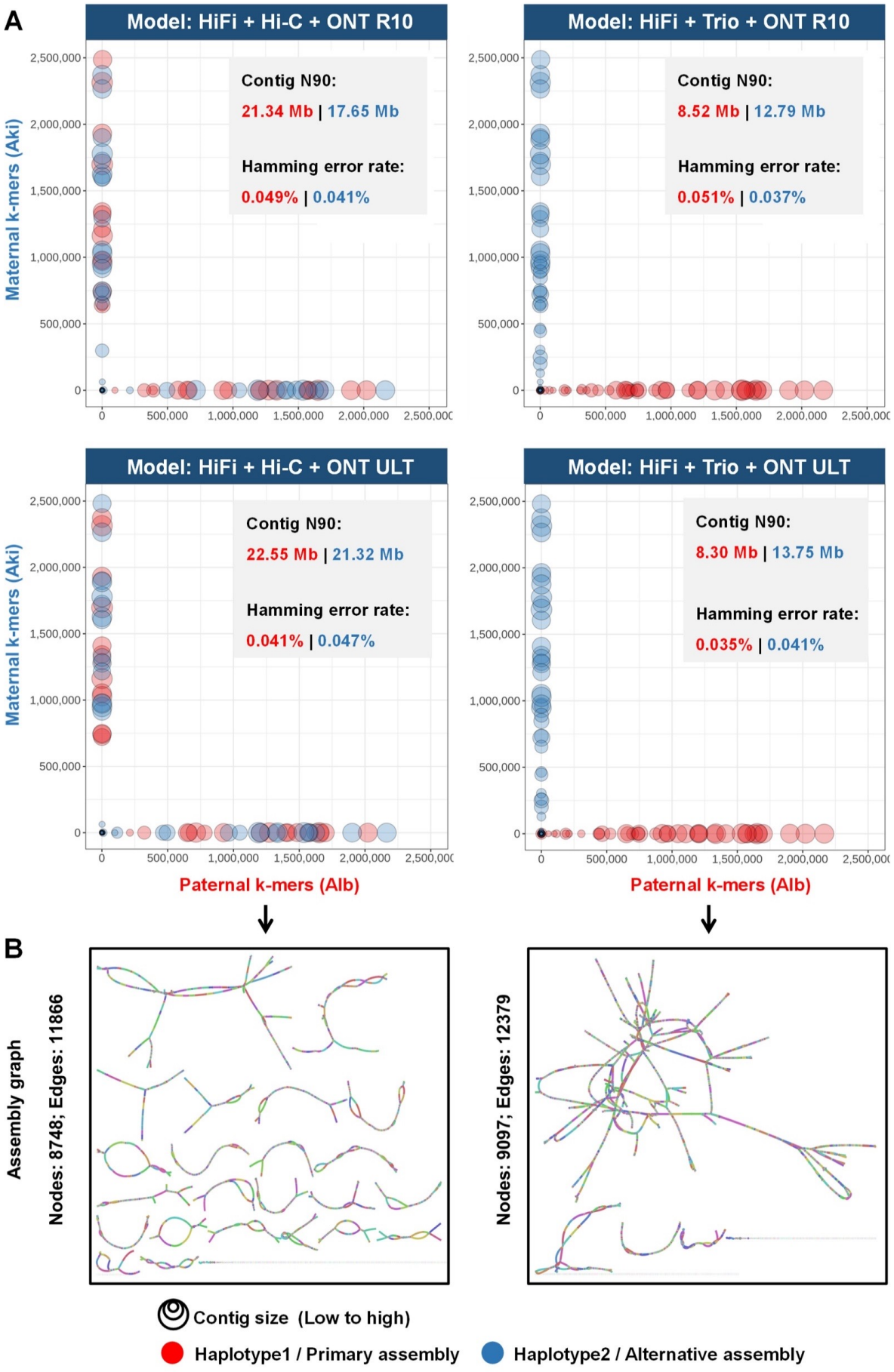


Fig. S4 | Effects of different sequencing depths on assembly quality. (A) The effect of different sequencing depths on assembly continuity. **(B)** The effect of different sequencing depths on phasing quality. **(C)** Comparison of phasing quality between sequencing depths of 10X and 40X. Note: Sequencing depth is calculated as the total number of bases from HiFi sequencing divided by 800 Mb. Red indicates the paternal haplotype, while blue denotes the maternal haplotype.

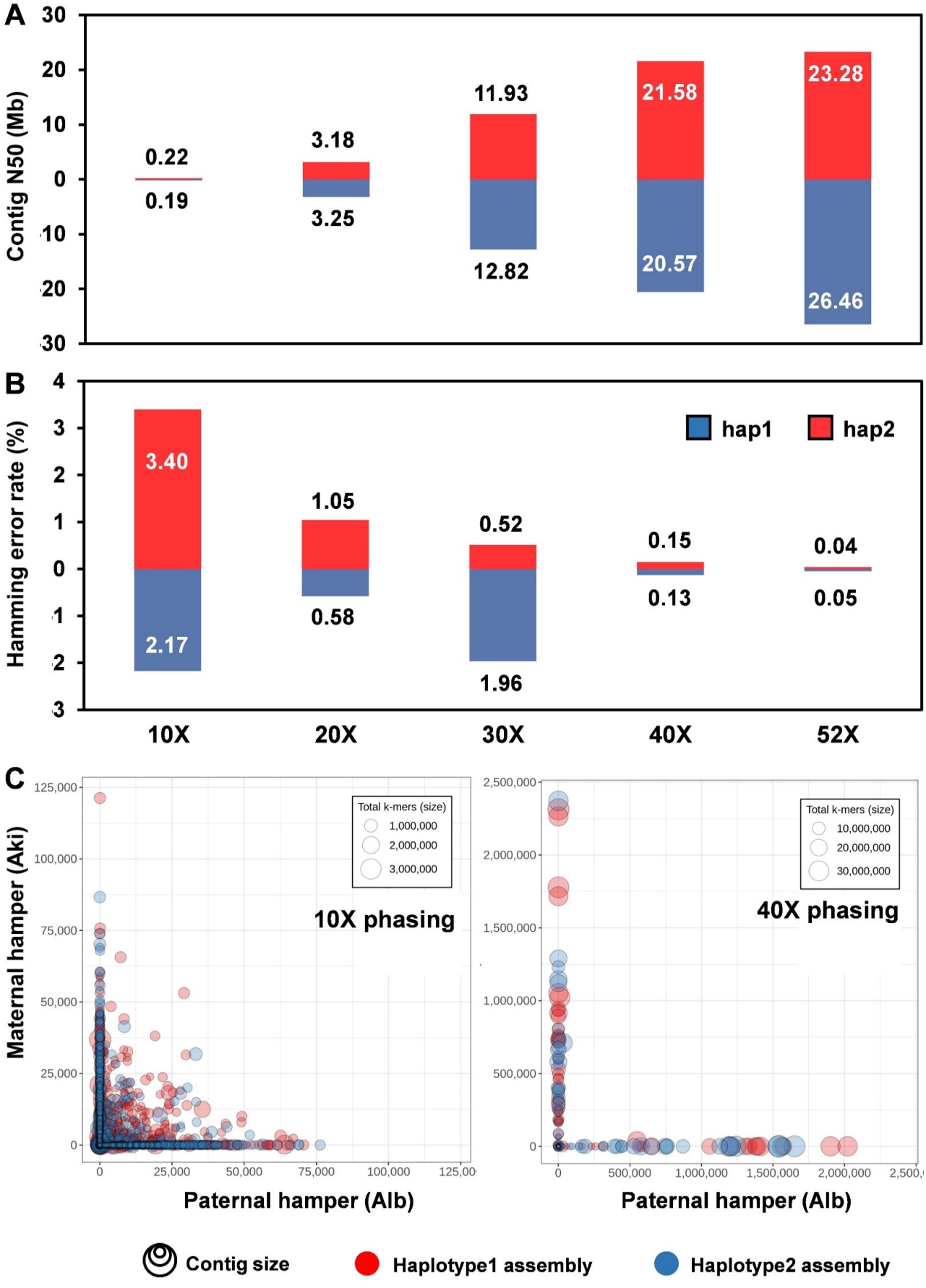
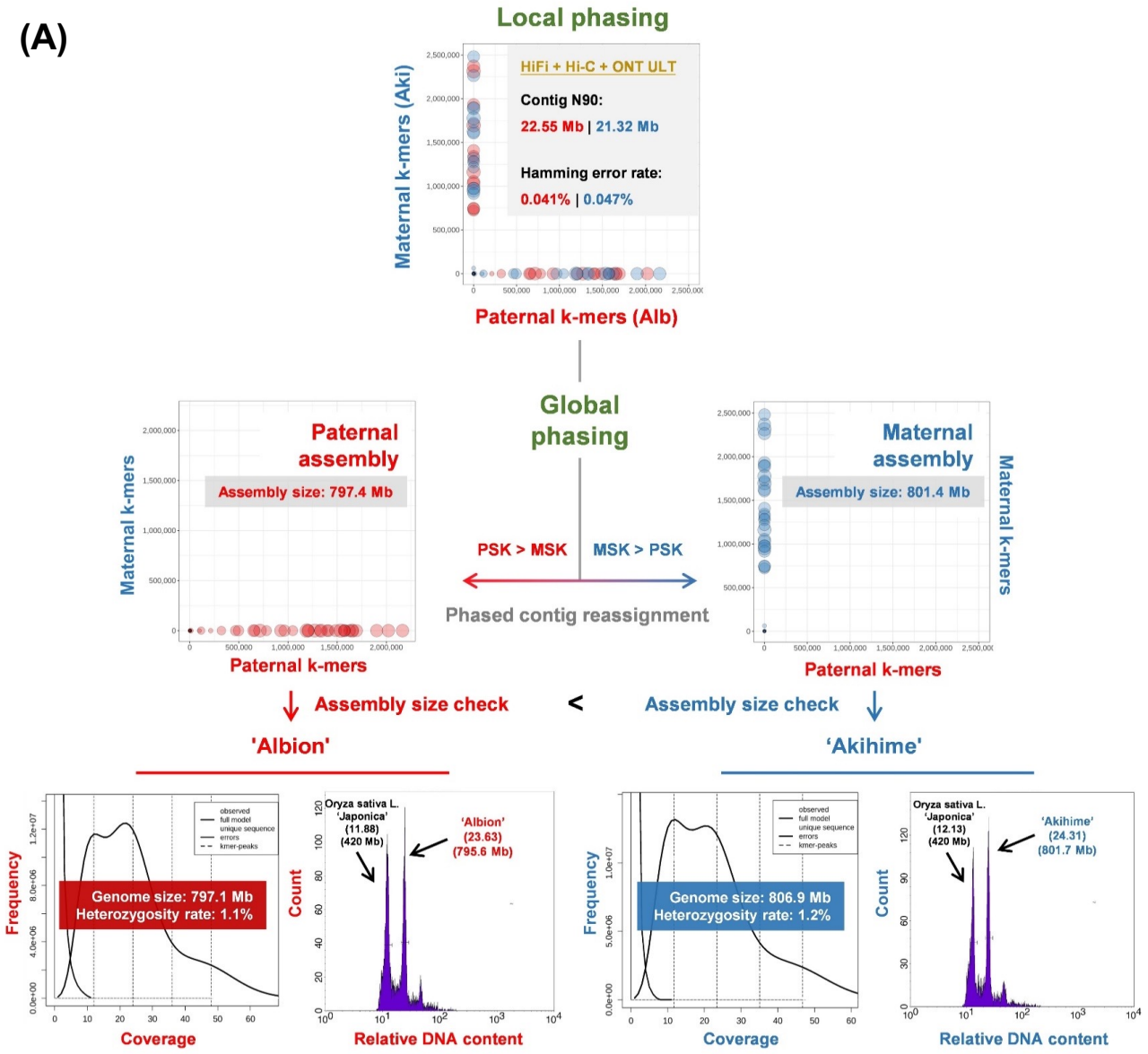


Fig. S5 | Two-step phased assembly strategy. (A) Comparison of assembled genome sizes of 'EA78' haplotypes against genome size estimates of its parental lines using k-mers and flow cytometry. Note: PSK represents paternal specific k-mers; Note: MSK represents maternal specific k-mers. **(B)** Putative limitations of the two-step phased assembly strategy. This panel illustrates the limitations encountered when contigs lack paternal-specific k-mers. The red lines represent contigs enriched with paternal-specific k-mers, the blue lines denote contigs with a high abundance of maternal-specific k-mers, and the grey lines indicates contigs (~0.074% of total assembled contigs) that does not contain any paternal-specific k-mers.



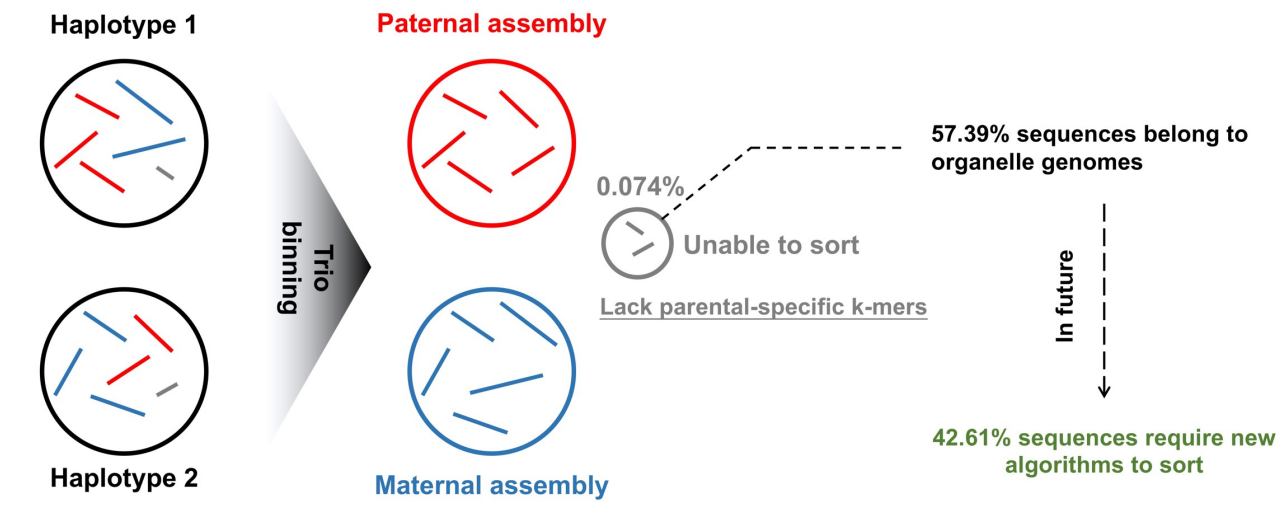


Fig. S6 | Evaluation of 'EA78' genome assembly quality. (A) Heat map of Hi-C interactions of the two haplotypes. High densities of interactions are indicated in red. **(B)** Macrosynteny between the diploid *F. vesca* reference genome (2n=2x=14) and the 'EA78' genome (2n=8x=56). **(C)** Genomic synteny between 'EA78' maternal (n=28) and paternal (n=28) haplotypes. **(D)** Identification of telomeric satellites in 'EA78' chromosomes. **(E)** Histogram of sequencing depth of HiFi reads mapped to the assemblies of the two haplotypes. Note: The major peak was shown with red line. **(F)** Merqury spectrum plots for kmer-based QV and completeness evaluation. **(G)** Assessment of the continuity of the 'EA78' genome using LTR assembly index (LAI).

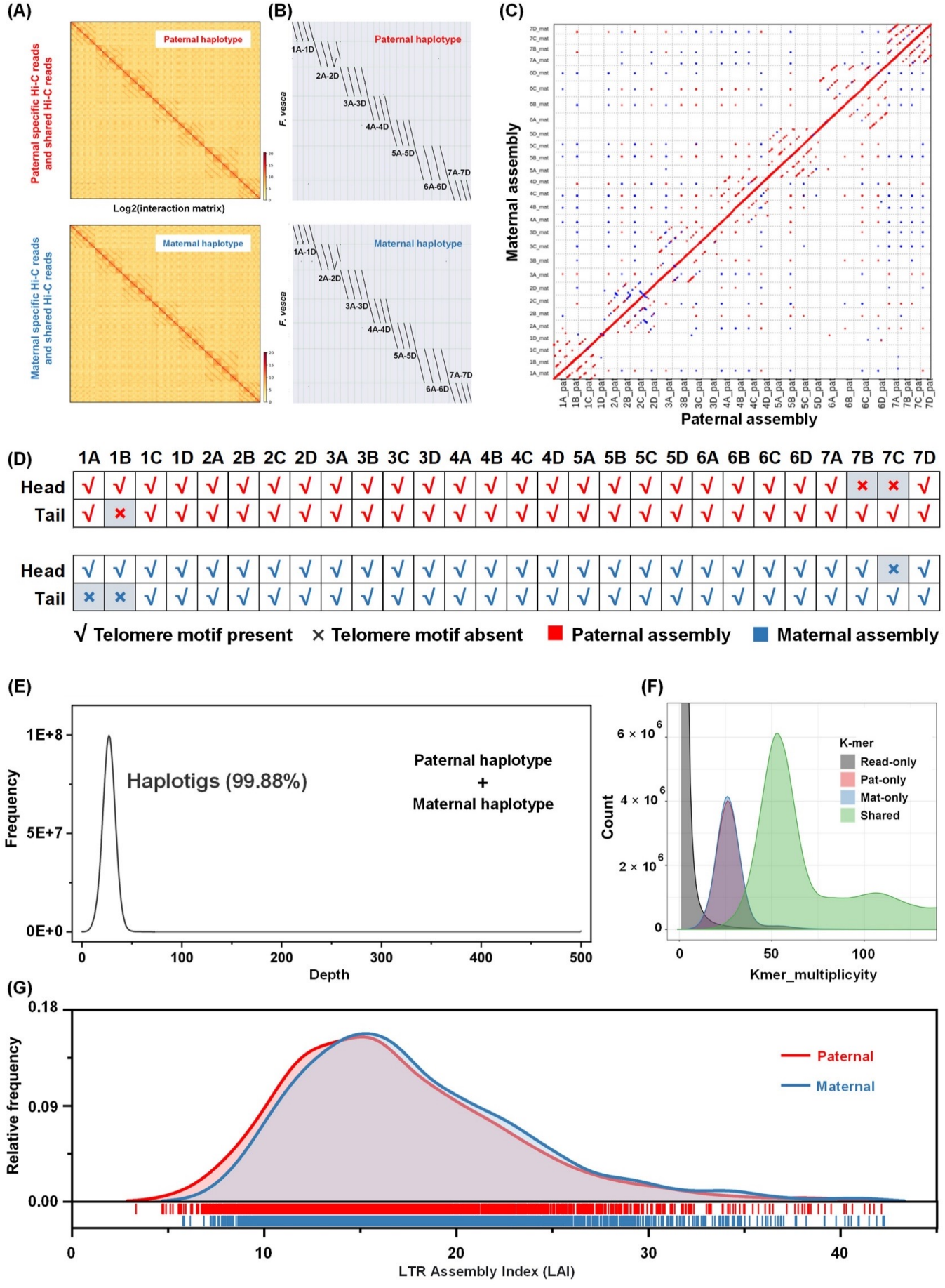


Fig. S7 | Mapping HiFi reads of American and Arian to the 'EA78' reference assemblies. Map of American (red font) and Asian (blue font) varieties HiFi reads on two haplotypes of 'EA78' respectively.

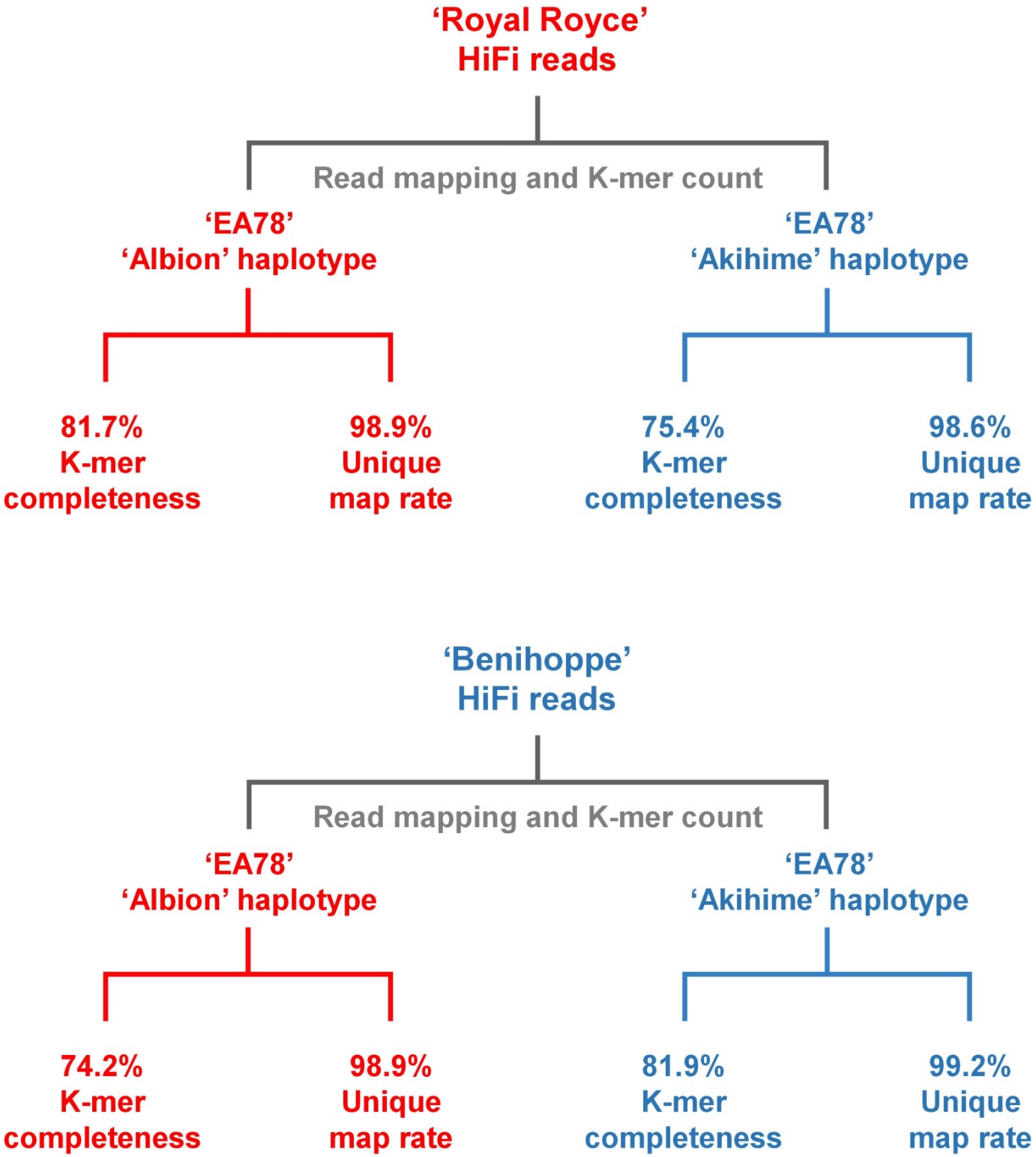


Fig. S8 | Identification of putative centromeres in 'EA78'. (A) Types and abundance of tandem repeats identified in the paternal and maternal assemblies, illustrating five shared clusters (CL) between the two haplotypes. This panel also highlights the similarity of *F. vesca* centromeric satellites to cluster 3 (CL3). Note: Average Nucleotide Identity (ANI). **(B)** Relative position and abundance of centromeric satellites across each chromosome of 'EA78'. High abundance of centromeric monomers is indicated in orange, while low abundance is shown in black. **(C)** FISH detection using the canonical centromeric satellite sequence as a probe for in *F. × ananassa* cv. 'Akihime' (Maternal material) (scale bar = 5 μm).

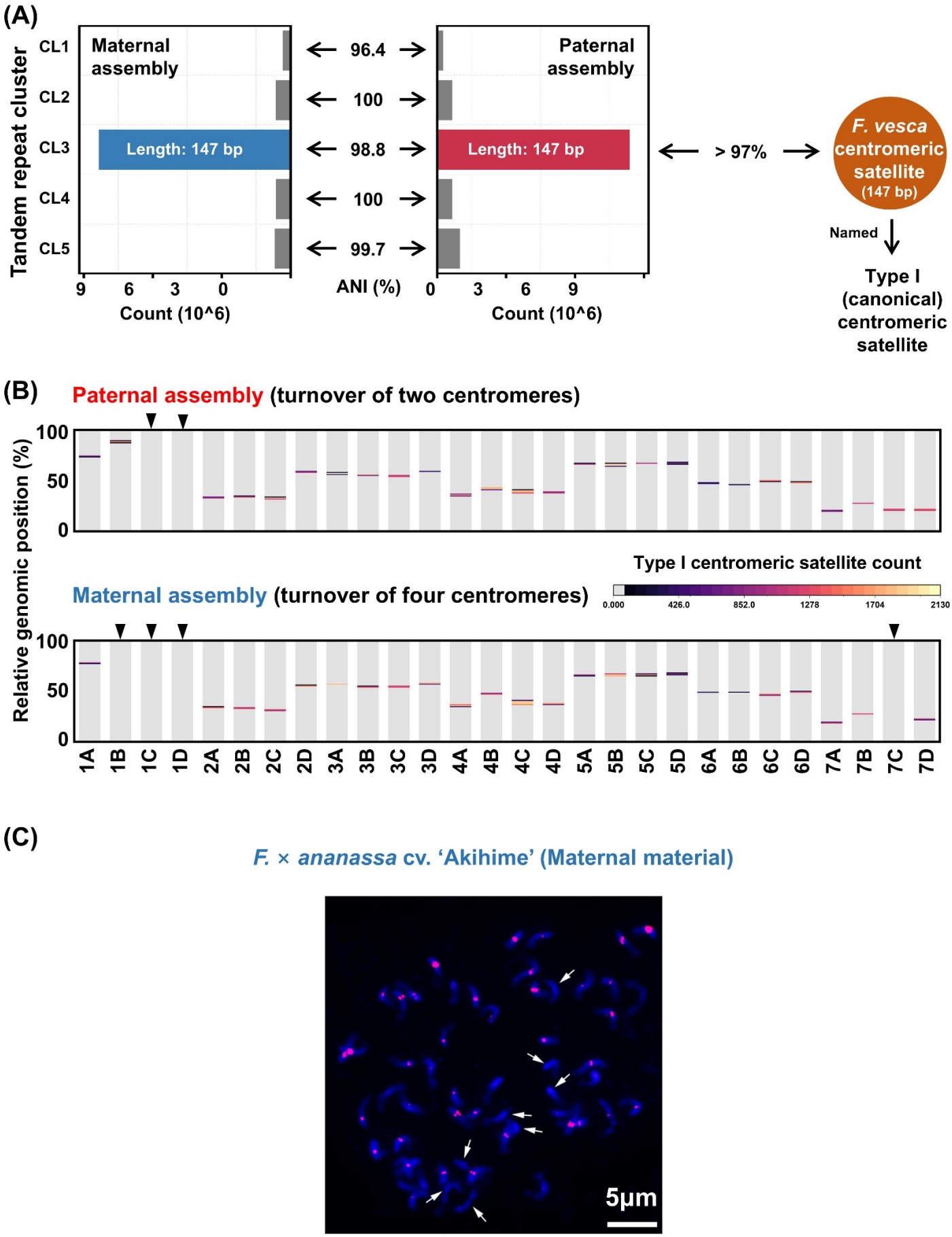
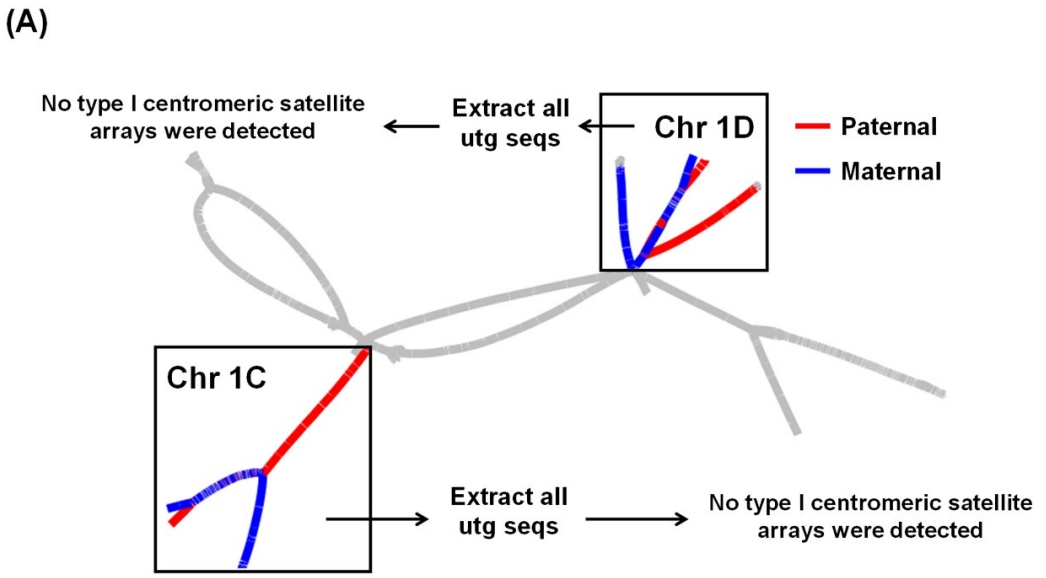
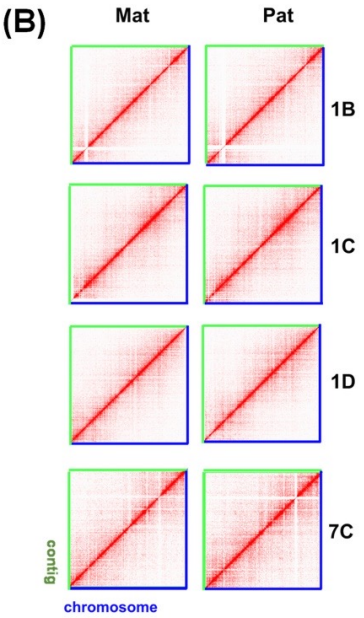


Fig. S9 | Validation of canonical centromeric satellite turnover on chromosomes 1B, 1C, 1D, and 7C. (A) Assembly graphs for 'EA78' chromosomes 1C and 1D, with red, blue, and grey lines representing sequences from the paternal haplotype, the maternal haplotypes, and other homoeologous chromosomes, respectively. **(B)** Hi-C interaction heatmaps for chromosome 1C and 1D, including both paternal (Pat) and maternal (Mat) assemblies. **(C)** Examination of canonical centromeric satellite turnover through comparisons of different assembly strategies and various sequencing data.





(C)

		Paternal			Maternal		
		1C	1D	1B	1C	1D	7C
HiFiasm	HiFi + Hi-C	×	×	×	×	×	×
	HiFi + Hi-C + R10	×	×	×	×	×	×
	HiFi + Hi-C + Ult	×	×	×	×	×	×
	HiFi + Trio + R10	×	×	×	×	×	×
	HiFi + Trio + Ult	×	×	×	×	×	×
Verkko	HiFi + Hi-C	×	×	×	×	×	×
	HiFi + Hi-C + R10	×	×	×	×	×	×
	HiFi + Hi-C + Ult	×	×	×	×	×	×
	HiFi + Trio + R10	×	×	×	×	×	×
	HiFi + Trio + Ult	×	×	×	×	×	×

× Type I centromeric satellite absent

Fig. S10 | Validation of the absence of canonical centromeric satellite in maternal chromosome 1B and its lack of correlation with phasing errors.

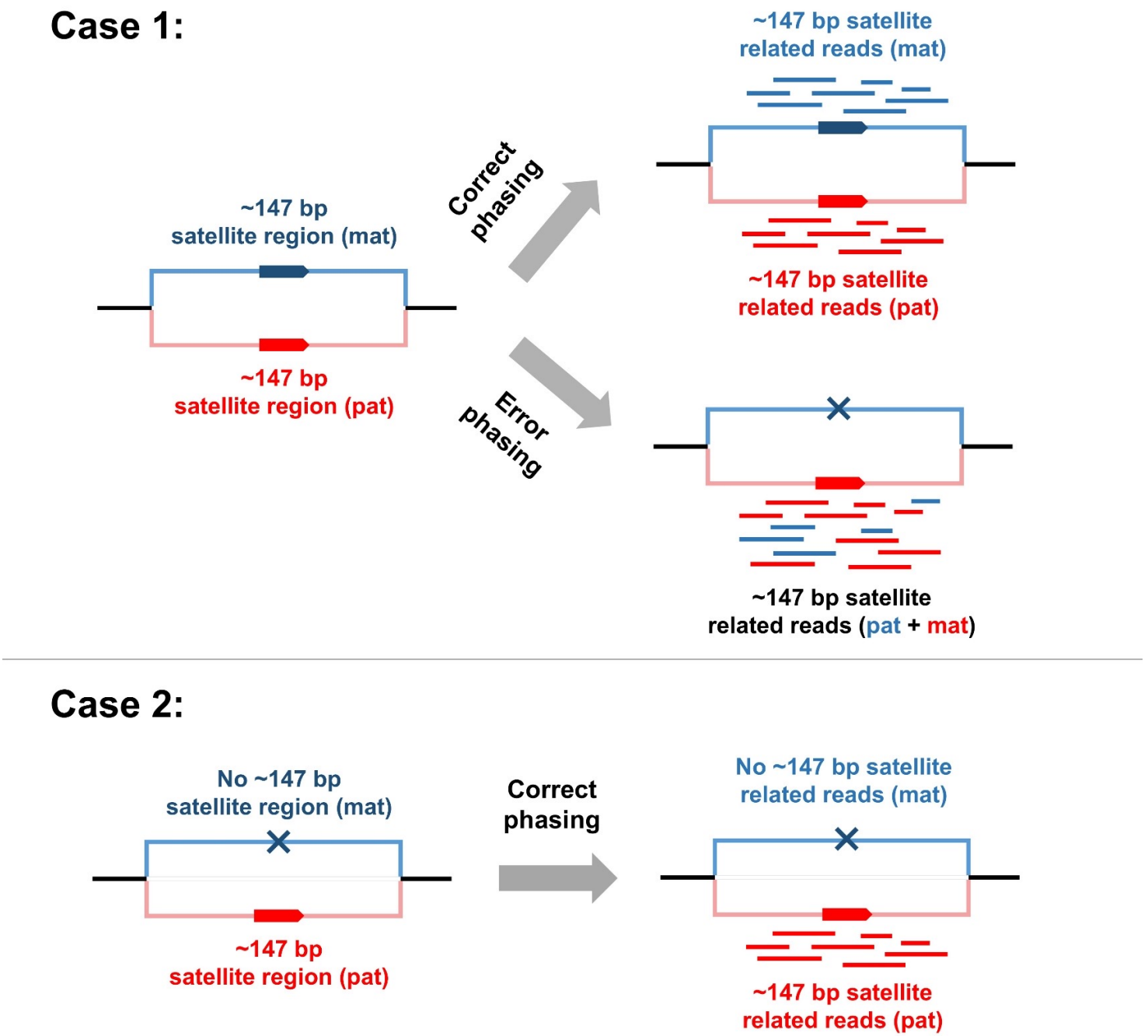


Fig. S11 | Inspection of CENH3-occupied region assembly. The assembly continuity and base accuracy of centromere regions are evaluated using HiFi reads.

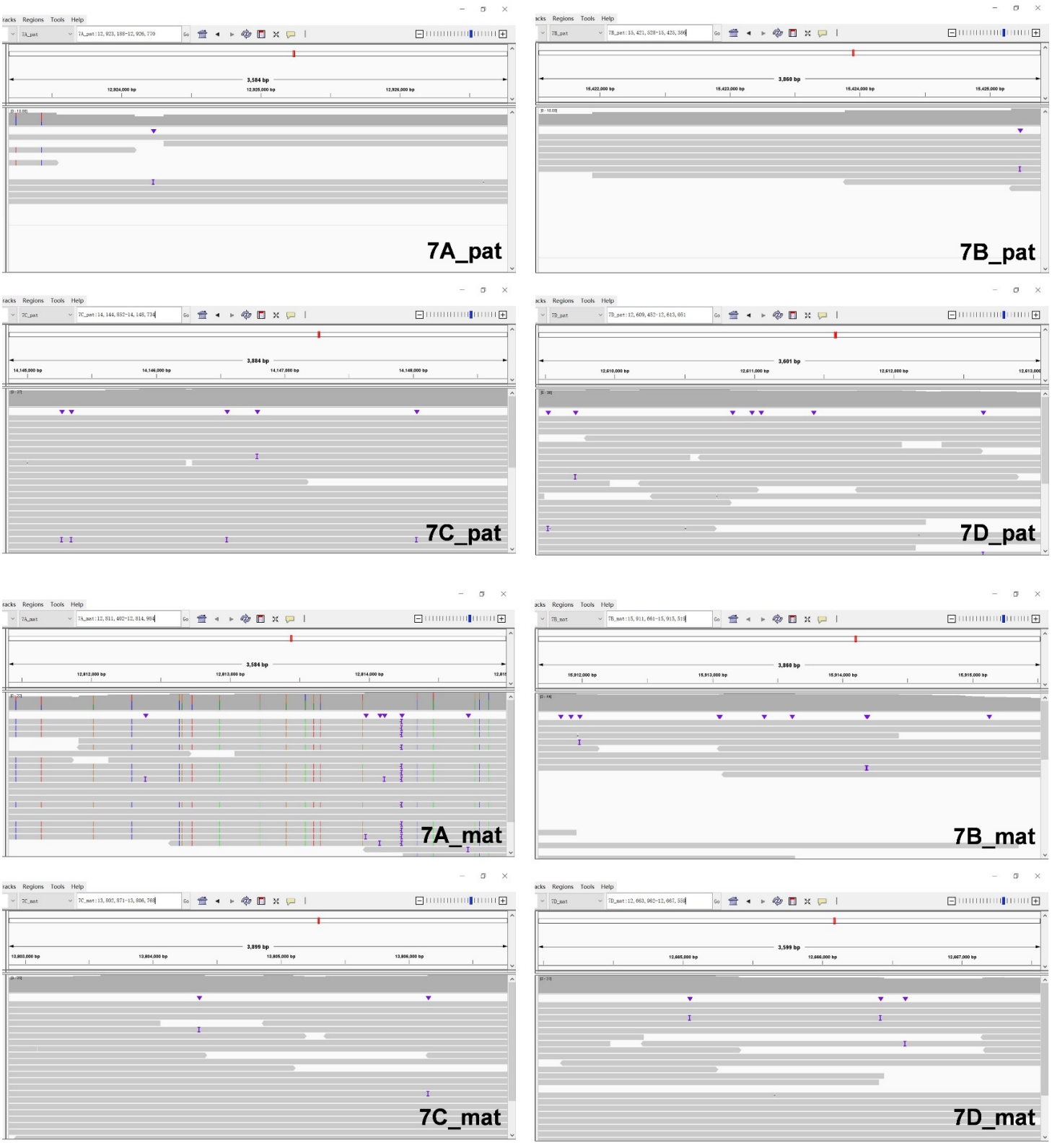


Fig. S12 | CEHN3 antigen survey and antibody production. (A) Location of CENH3 on the genome (shown as the eight homologous chromosome of chr7). **(B)** Hydrophilicity analysis of the antigen. **(C)** Epitope exposure analysis of the antigen. **(D)** Immunogenicity analysis of the antigen. **(E)** Enzyme Linked Immunosorbent Assay (ELISA) serum titer assay. Note: K represents the dilution factor.

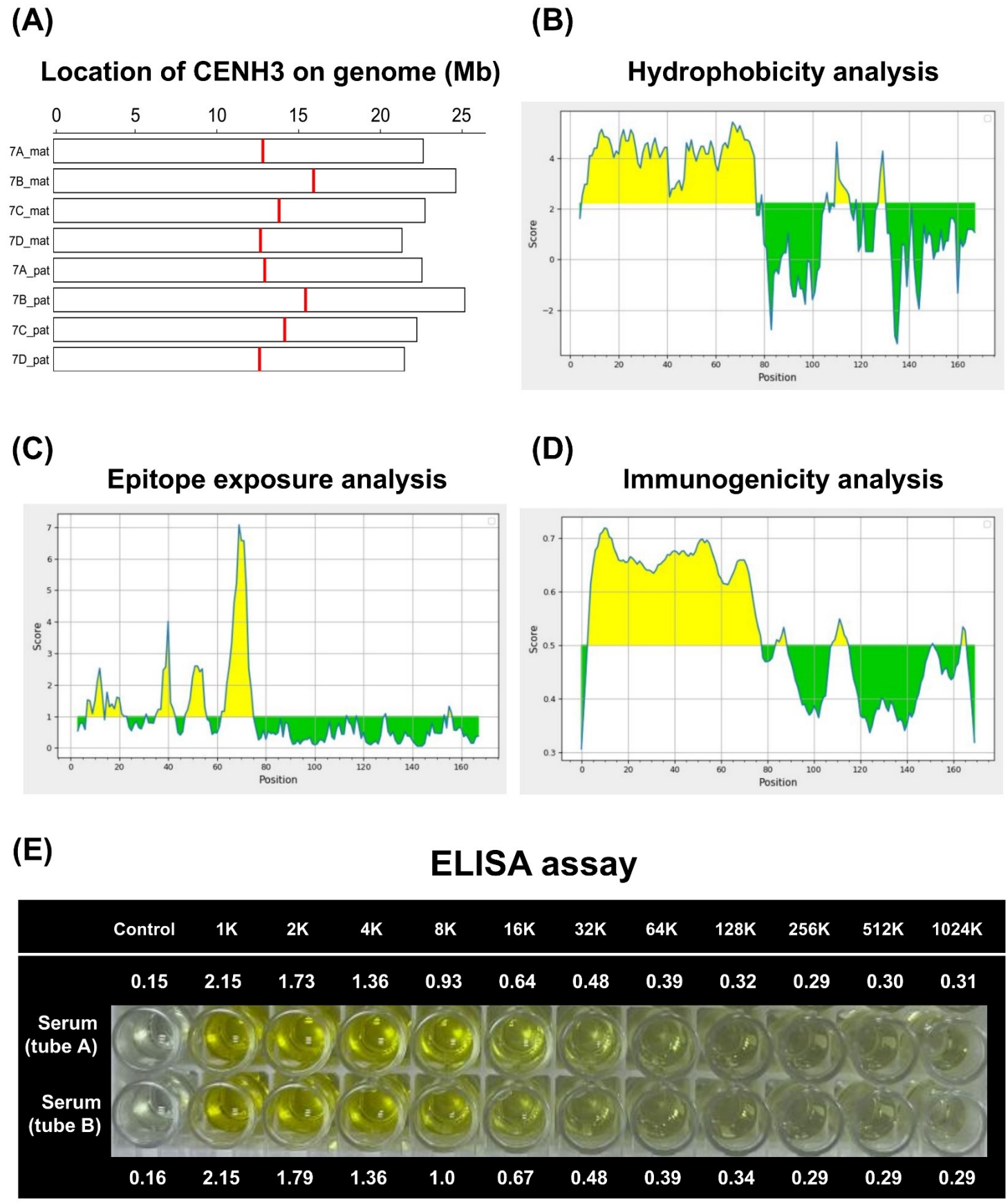


Fig. S13 | Identification of centromere regions in the paternal assembly of 'EA78'. Centromere regions were identified by integrating data from centromeric satellite analysis and CENH3-ChIP-seq. Bin size: 1000 bp.

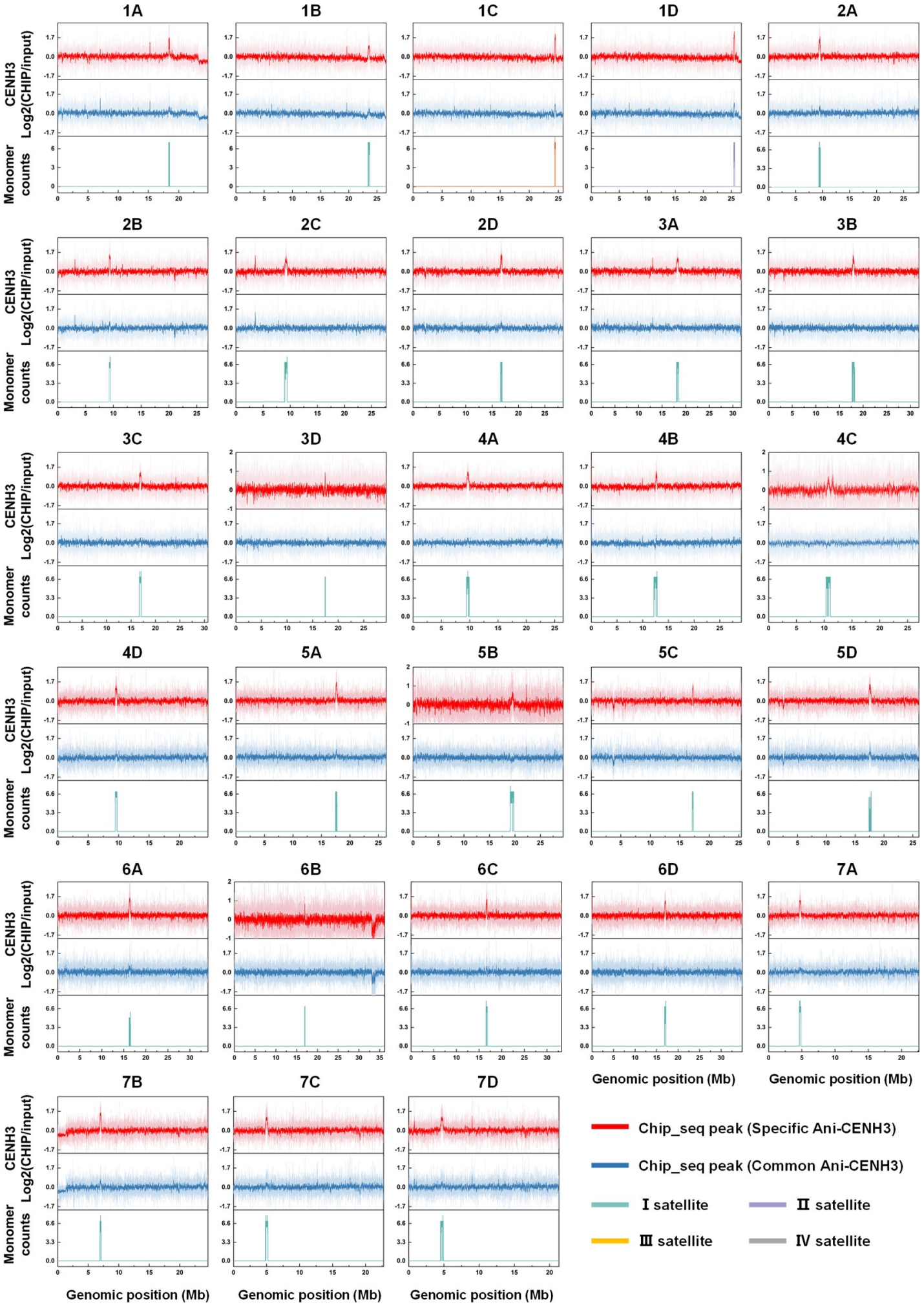


Fig. S14 | Identification of centromere regions in the maternal assembly of 'EA78'. Centromere regions were identified by integrating data from centromeric satellite analysis and CENH3-ChIP-seq. Bin size: 1000 bp.

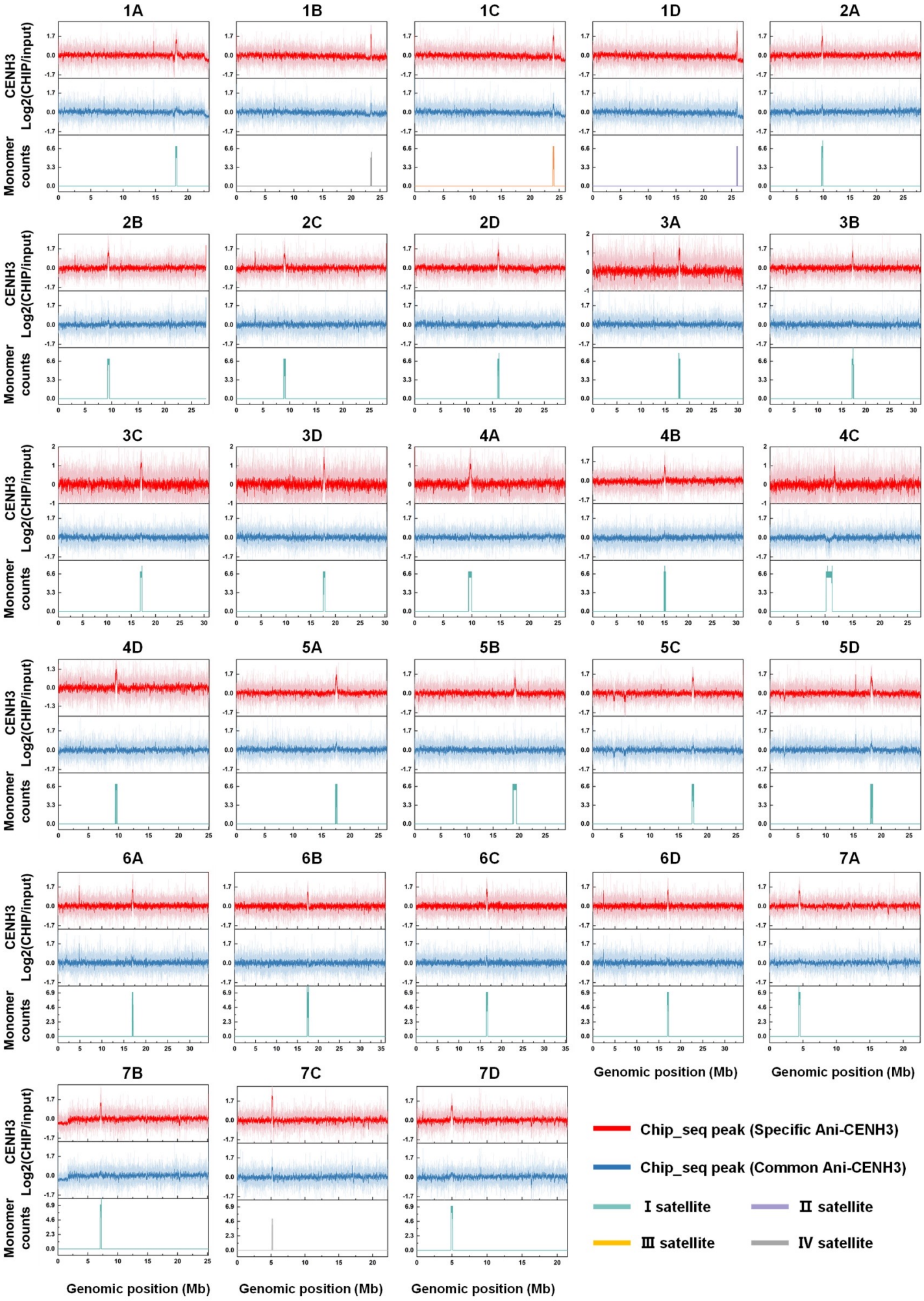


Fig. S15 | Centromere repositioning. (A) Identification of centromeric regions in the maternal assembly by integrating centromeric satellite analysis and CENH3 ChIP enrichment. Bin size: 1000 bp. **(B)** Structural features of regions enriched ChIP-seq signals and centromeric satellites. Bin size: 2000 bp.

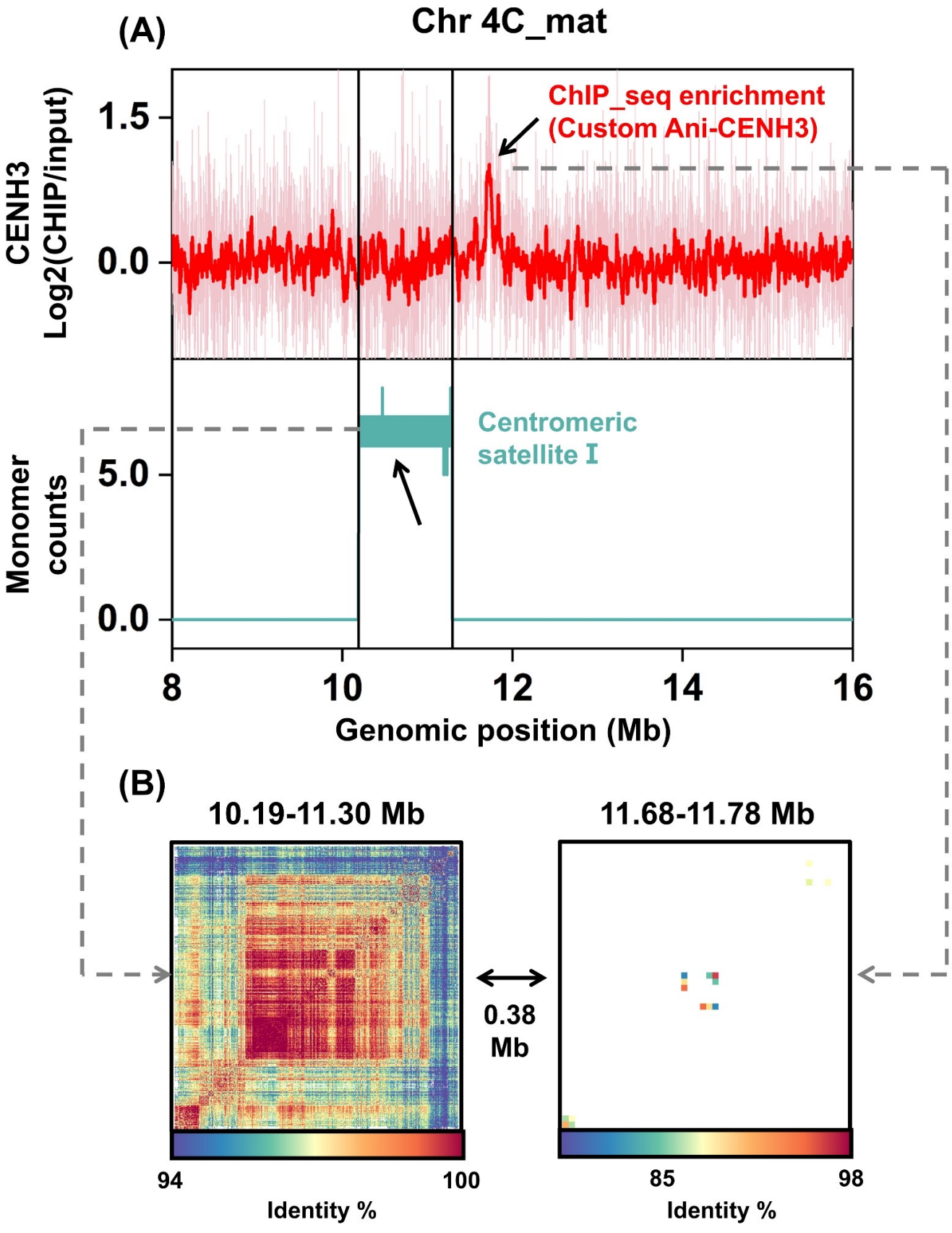


Fig. S16 | Genomic features of 'EA78' centromeres. **(A)** Centromere length across chromosomes. **(B)** Correlation between centromere length and the number of centromeric monomers. Note: Different colors represent subgenomes of 'EA78': red for A, blue for B, orange for C, and cyan for D. Circles correspond to chromosomes of the paternal haplotype, while triangles represent the maternal haplotype. **(C)** Centromeric Long Terminal Repeat (LTR) similarity matrix, based on Average Nucleotide Identity (ANI). **(D)** Methylation state of centromeric regions in 'EA78' chromosomes. The levels of three methylation types (CG, CHH, CHG) across each centromere region are shown.

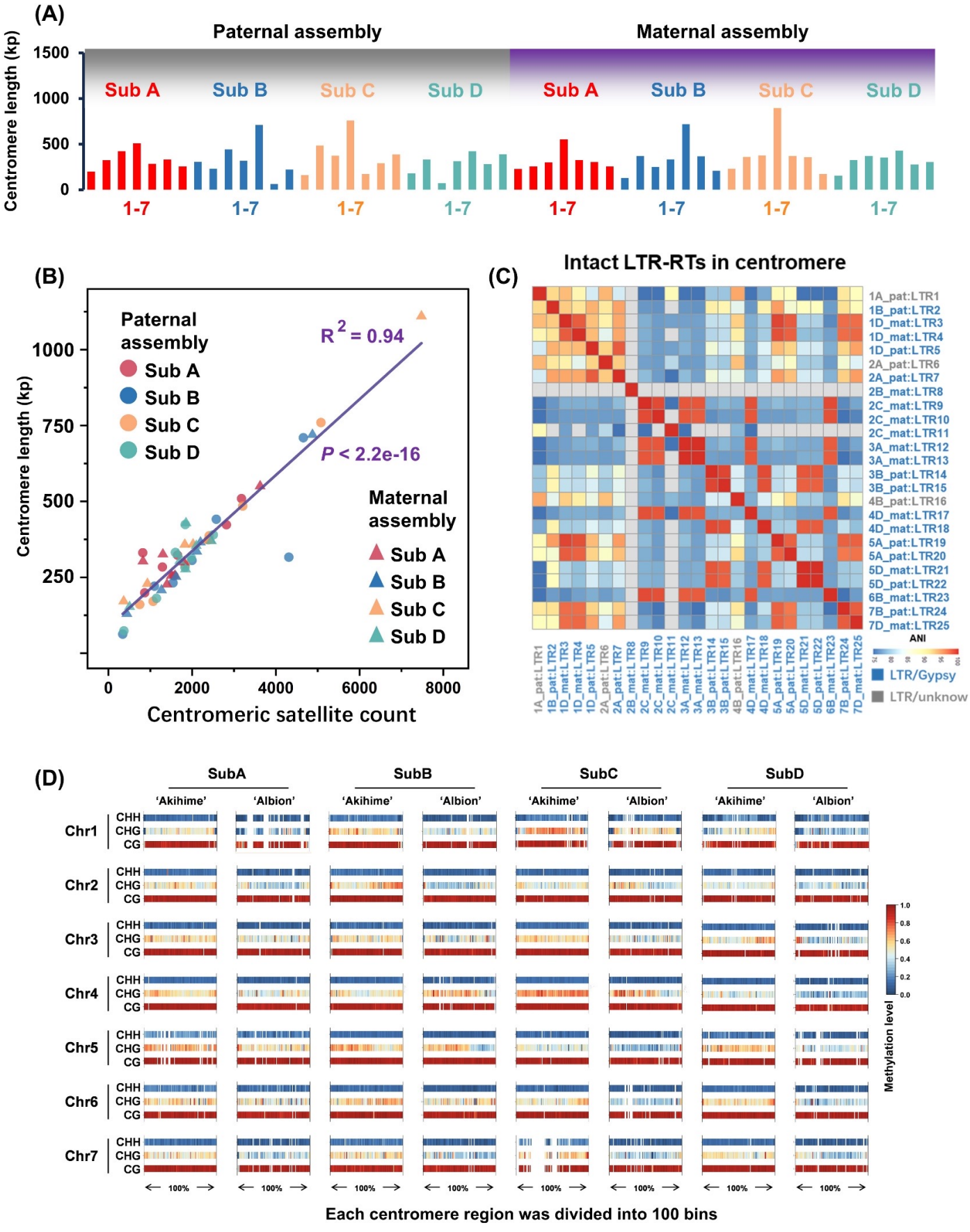


Fig. S17 | Sequence similarity and divergence of four centromeric satellite types (A) Centromere similarity matrix showing Average Nucleotide Identity (ANI). Note: for sequences with unidentified ANI values, NA was replaced with a similarity of <75%. **(B)** Principal Coordinate Analysis (PCoA) distribution of the four types of centromeric satellites, using average nucleotide similarity distance as input. Note: Average nucleotide similarity distance is calculated as 100 – ANI.

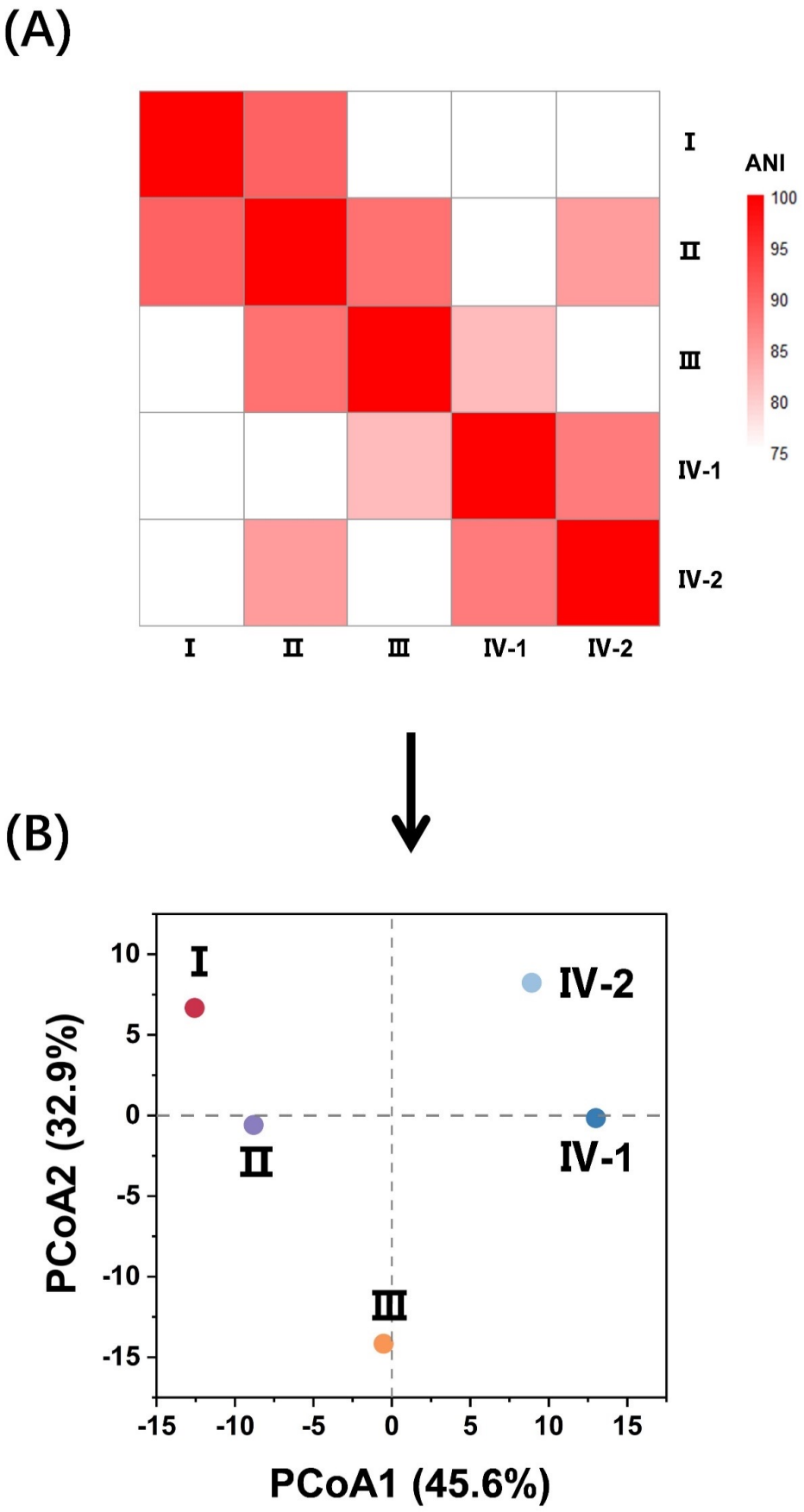


Fig. S18 | Sequence similarity of centromeric satellites among 'EA78' subgenomes. Principal Coordinate Analysis (PCoA) distribution of centromeric satellite sequences, using average nucleotide similarity distance as input. Note: Average nucleotide similarity distance is calculated as 100% – Average Nucleotide Identity (ANI) in percentage.

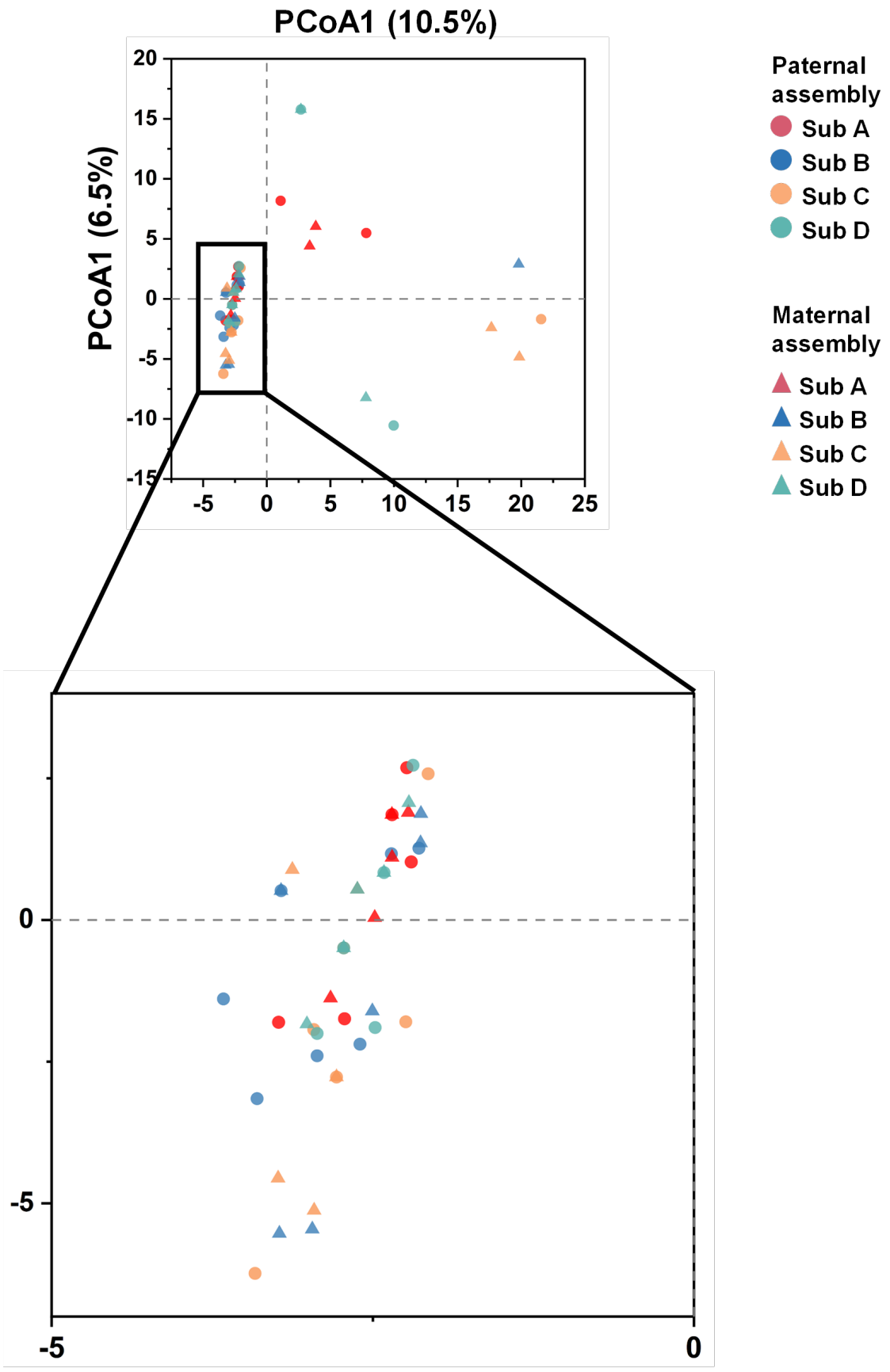


Fig. S19 | Comparison of the main satellite types across chromosomes among Asian, Asian and European-American, European-American varieties. Note: The squares of different colors represent different main satellite types. Different color fonts represent different varieties, with blue representing Asian variety, red representing European and American variety, and purple representing Asian-European hybrids.

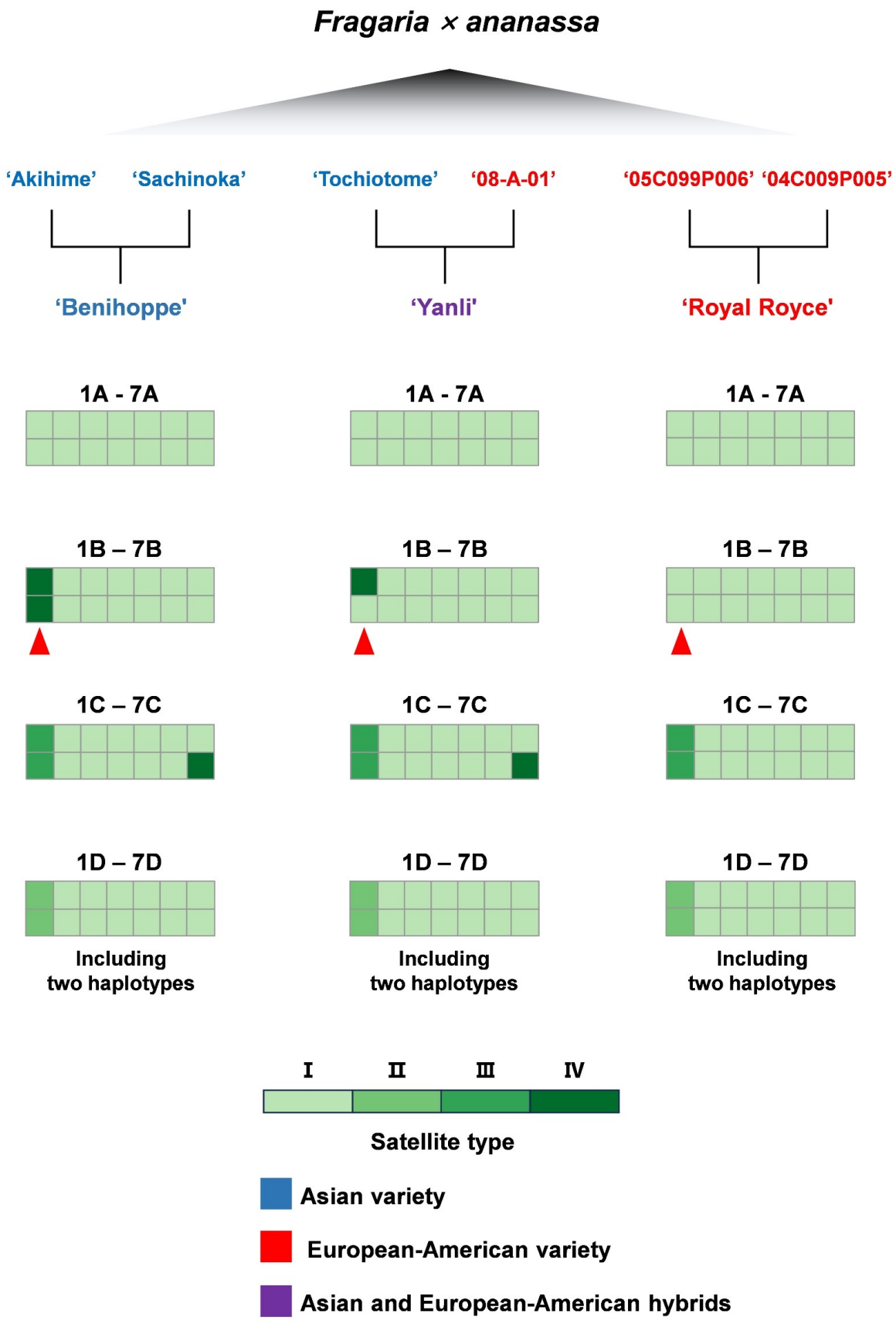


Fig. S20 | Simplified evolutionary history of hybridization and domestication of cultivated strawberry. Note: Different colors represent different subgenomes: red for the A subgenome, blue for the B subgenome, yellow for the C subgenome, and green for the D subgenome.

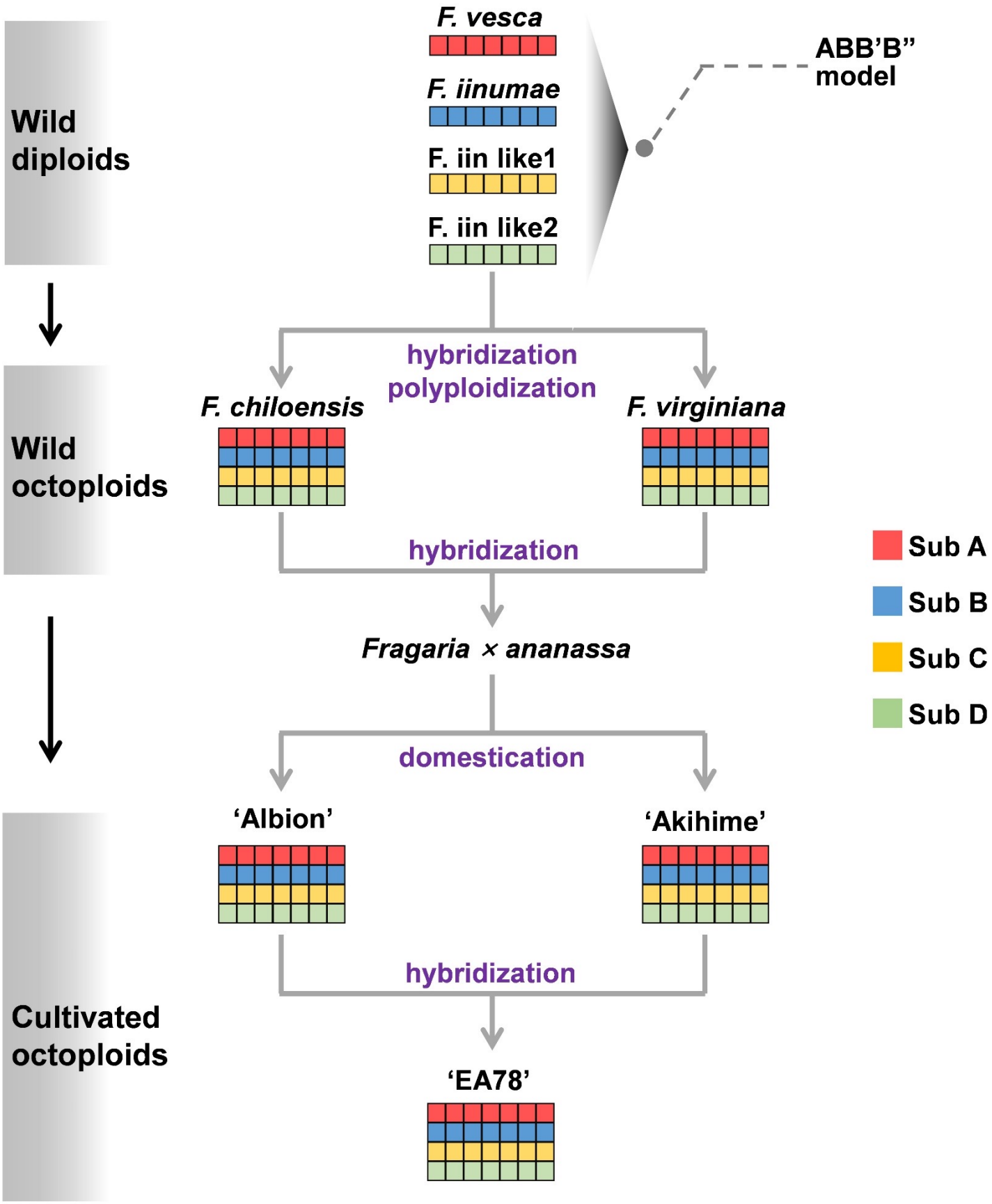
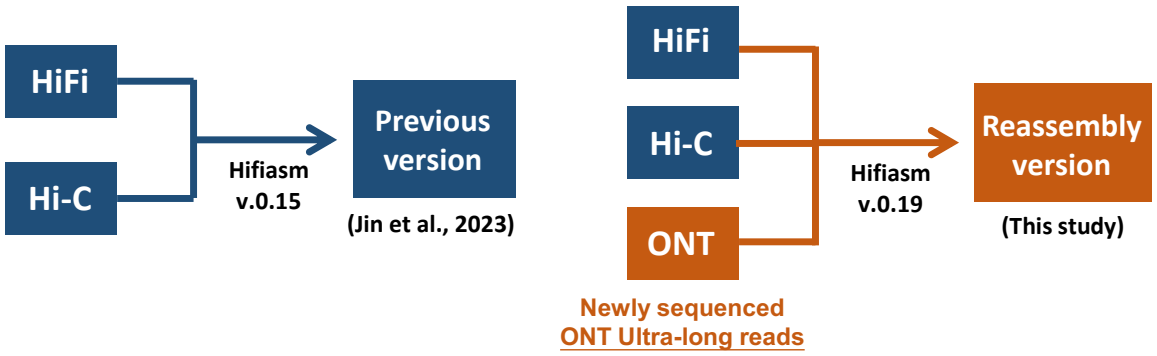


Fig. S21 | Reassembly of the two wild octoploid progenitors (*F. chiloensis* and *F. virginiana*) to improve genome continuity. **(A)** Comparison of genome assembly strategies used for assembling the two wild progenitors. **(B)** Comparison of genome assembly continuity between the previous and the reassembly versions. **(C)** Analysis of the overlap between gaps and centromere positions. Note: Triangles represent gap locations, and circles denote centromere positions.

(A)



(B)

	F. chiloensis		*F. virginiana*	
	Previous version (hap1 | hap2)	Reassembly version (hap1 | hap2)	Previous version (hap1 | hap2)	Reassembly version (hap1 | hap2)
Assembly szie (Mb)	839.9 | 824.2	806.4 | 833.8	787.8 | 769.2	771.1 | 771.8
Contig N50 (Mb)	11.3 | 8.9	17.1 | 17.9	14.3 | 14.2	16.3 | 16.2
Number of contigs	1305 | 632	391 | 241	995 | 531	318 | 209
Number of gaps	129 | 136	0 | 0	110 | 83	1 | 2

(C)

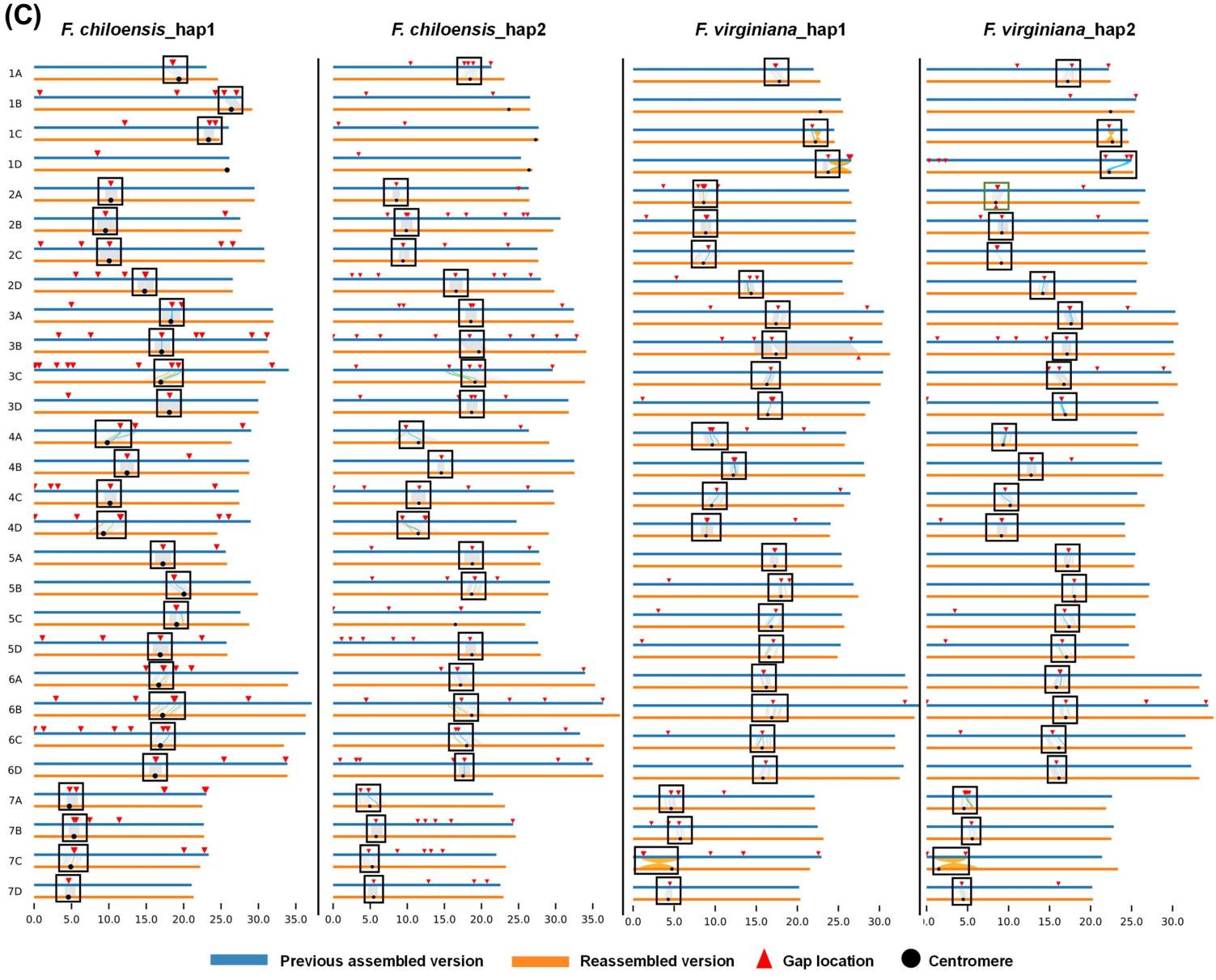


Fig. S22 | Heat map of Hi-C interactions of the two updated haplotype assemblies of *F. chiloensis* and *F. virginiana*.

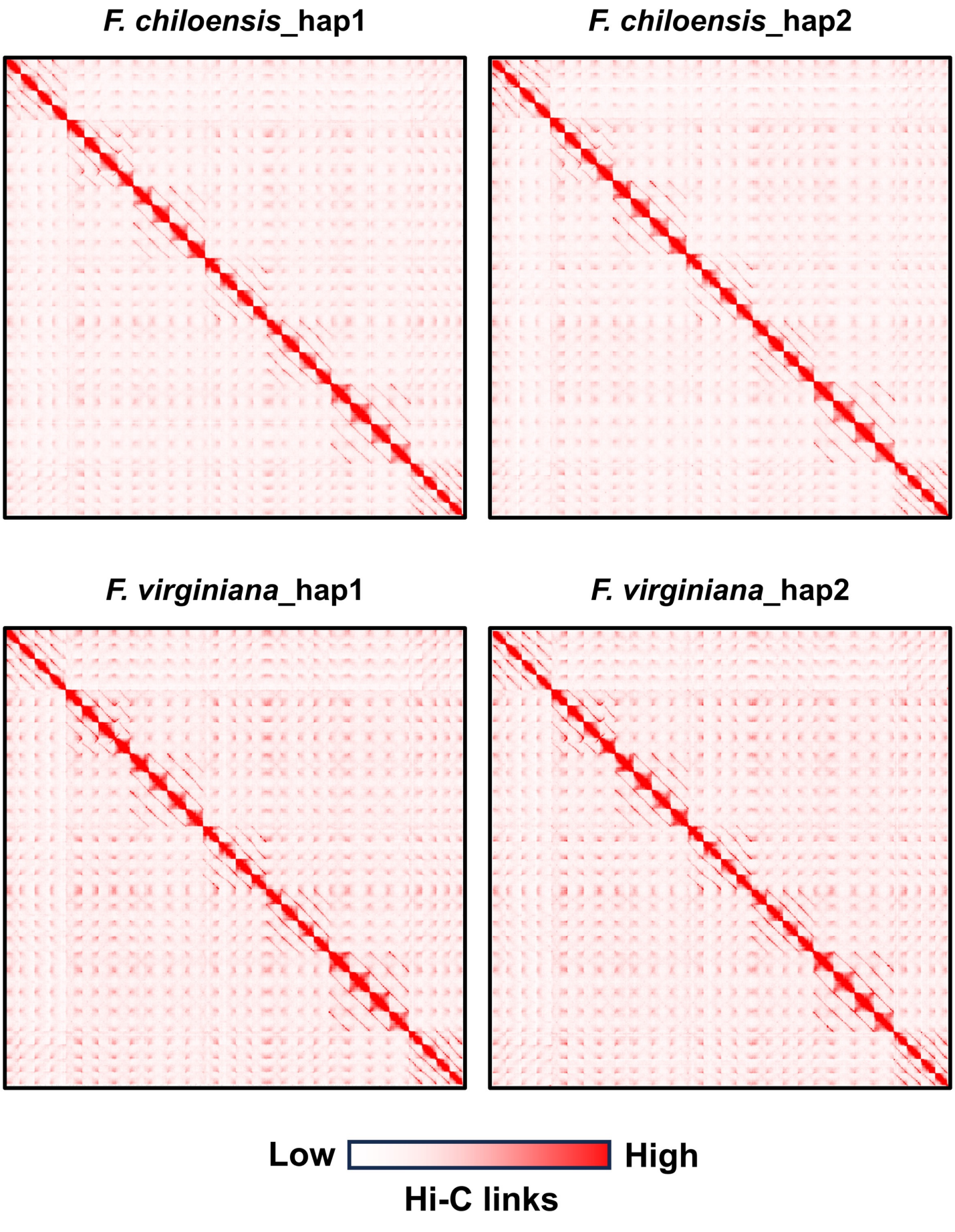


Fig. S23 | Centromere structural features of the wild octoploid strawberry *F. chiloensis*. Heatmaps shows the high-resolution sequence similarity matrix of centromere candidate regions (Bin size: 2000 bp).

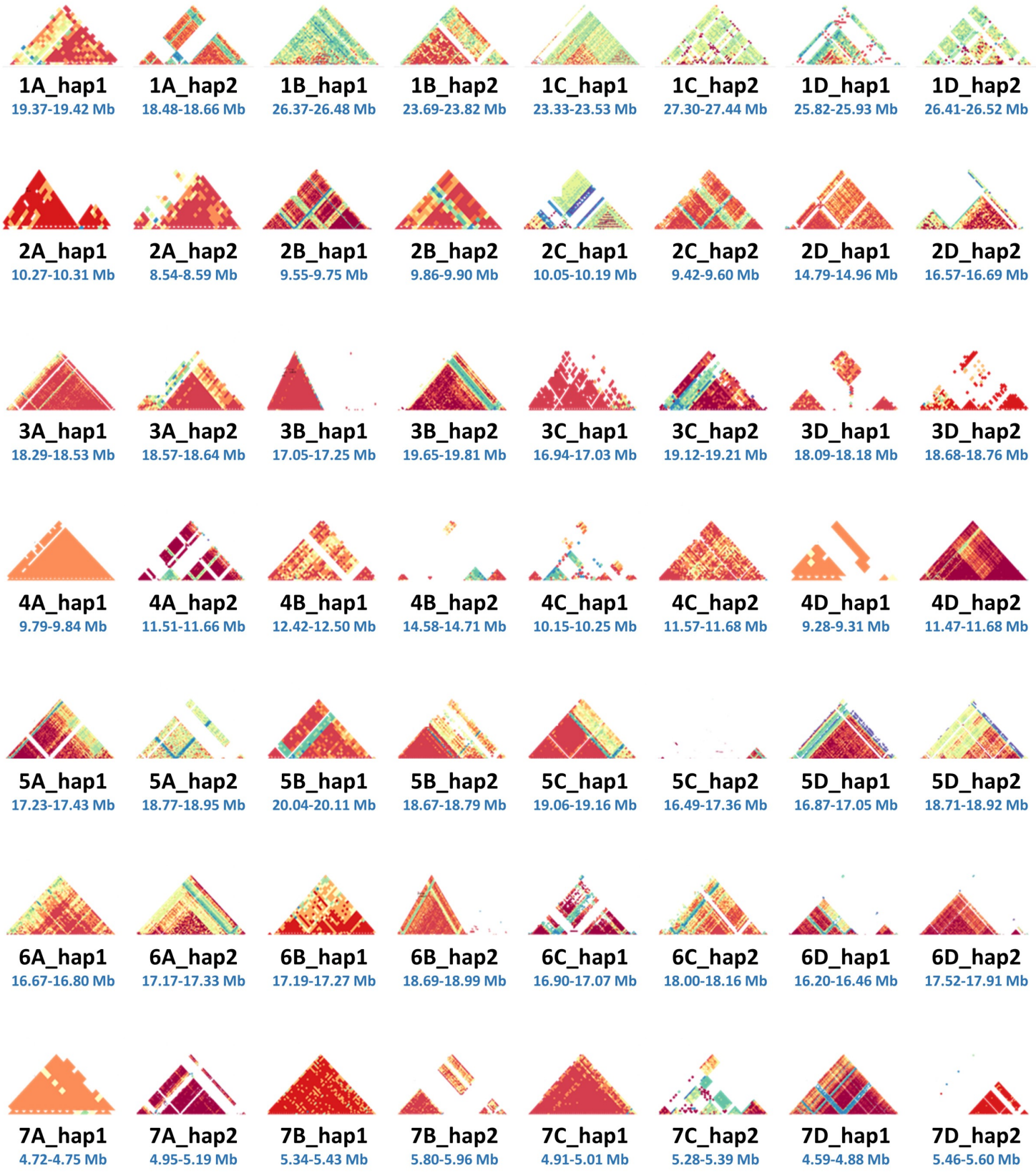


Fig. S24 | Centromere structural features of the wild octoploid strawberry *F. virginiana*. Heatmaps shows the high-resolution sequence similarity matrix of centromere candidate regions (Bin size: 2000 bp).

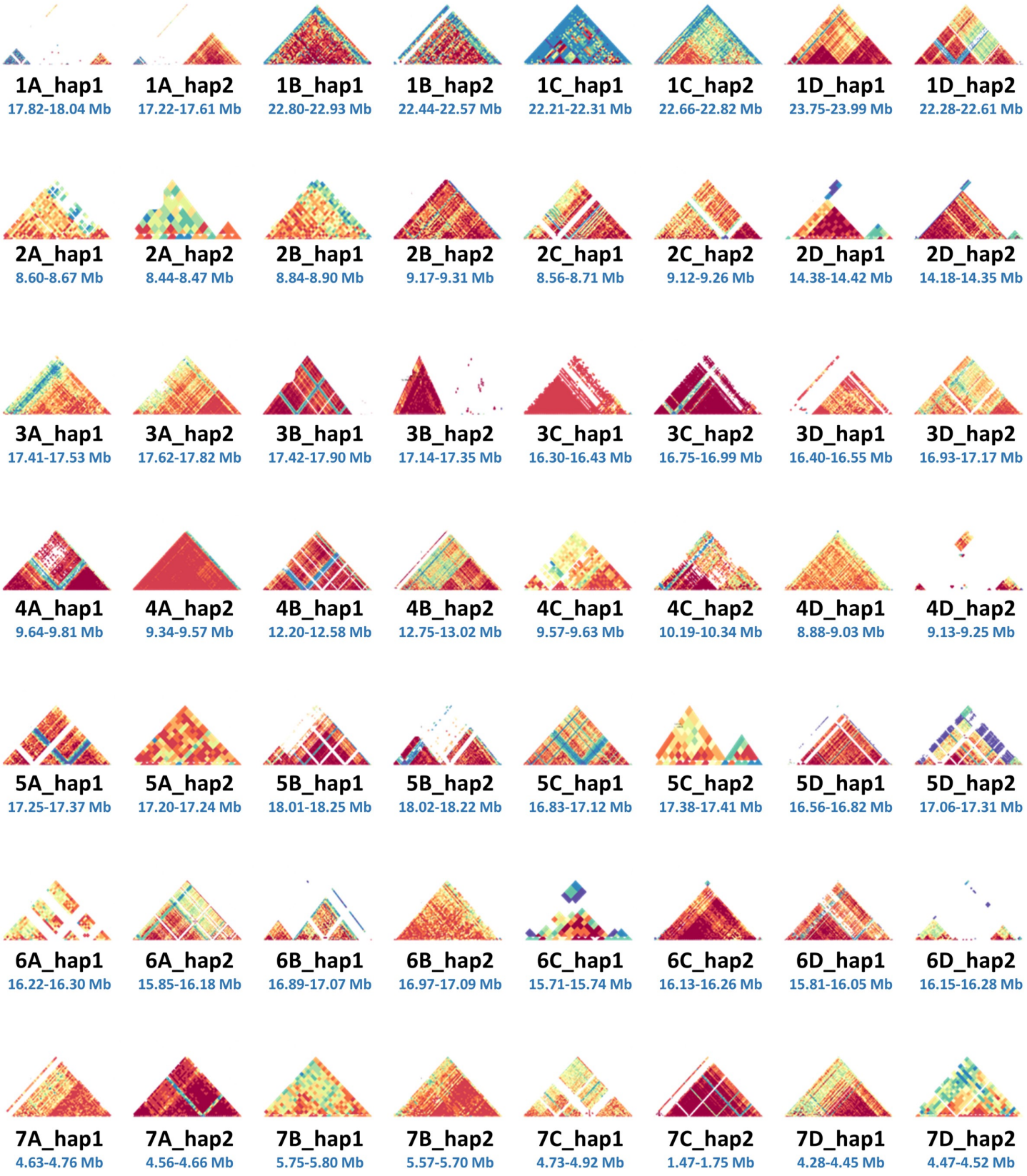


Fig. S25 | Centromere structural features of the cultivated strawberry 'EA78'. Heatmaps shows the high-resolution sequence similarity matrix of centromere candidate regions (Bin size: 2000 bp).

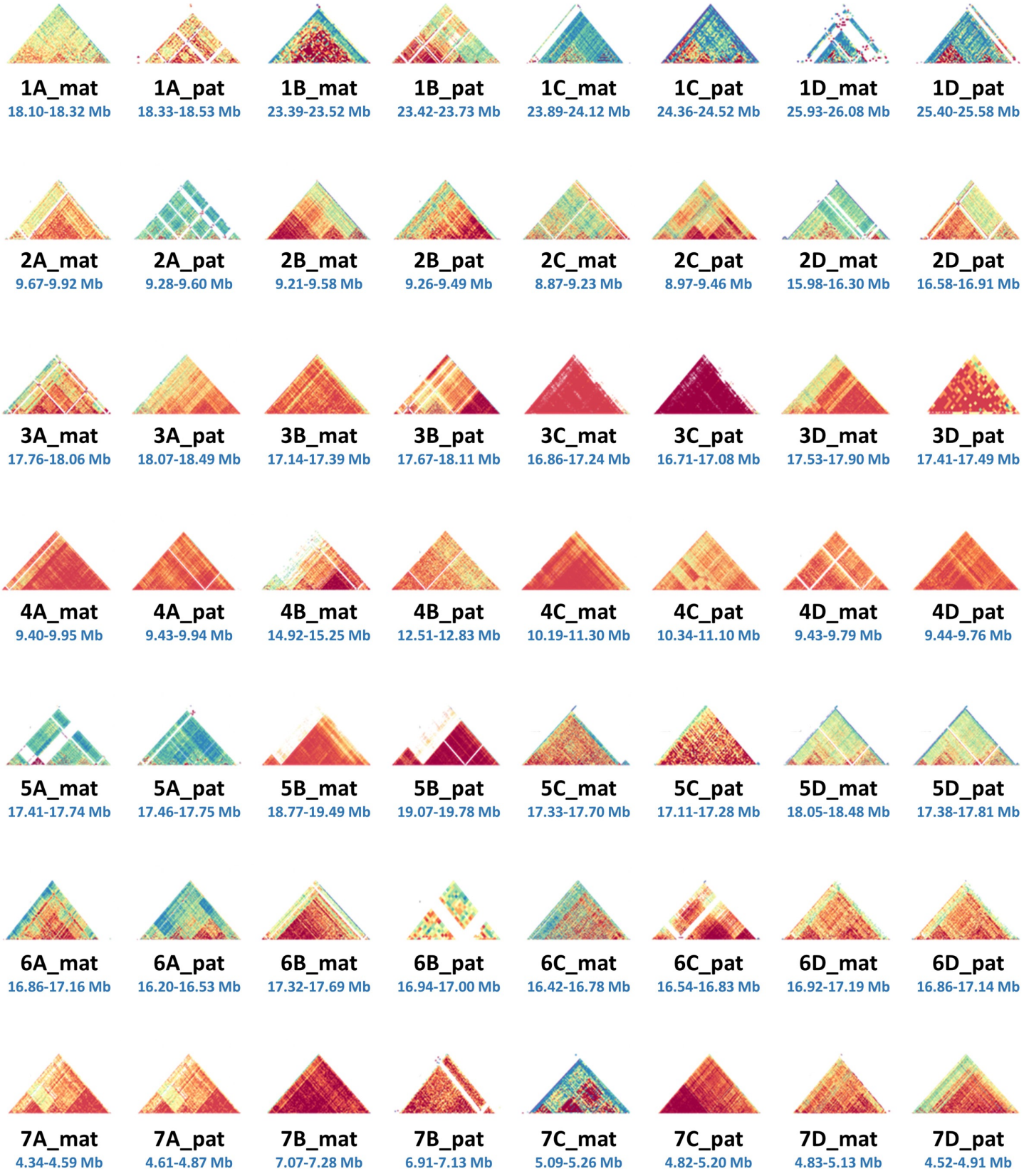


Fig. S26 | Centromere length variations among wild and cultivated strawberry species. (A) Comparison of centromere lengths between wild and cultivated species. Note: Both haplotypes are available for octoploid species (n=14), while only one haplotype is available for diploid species (n=7); Statistical analysis was performed using one-way ANOVA with Duncan's multiple range test; $P < 0.05$. **(B)** HiFi reads coverage across different species on the cultivated strawberry 'EA78' genome. The red, orange, purple, and blue lines represent the coverage of HiFi reads on 'EA78' genome for *F. vesca*, *F. virginiana*, *F. chiloensis*, and 'EA78', respectively.

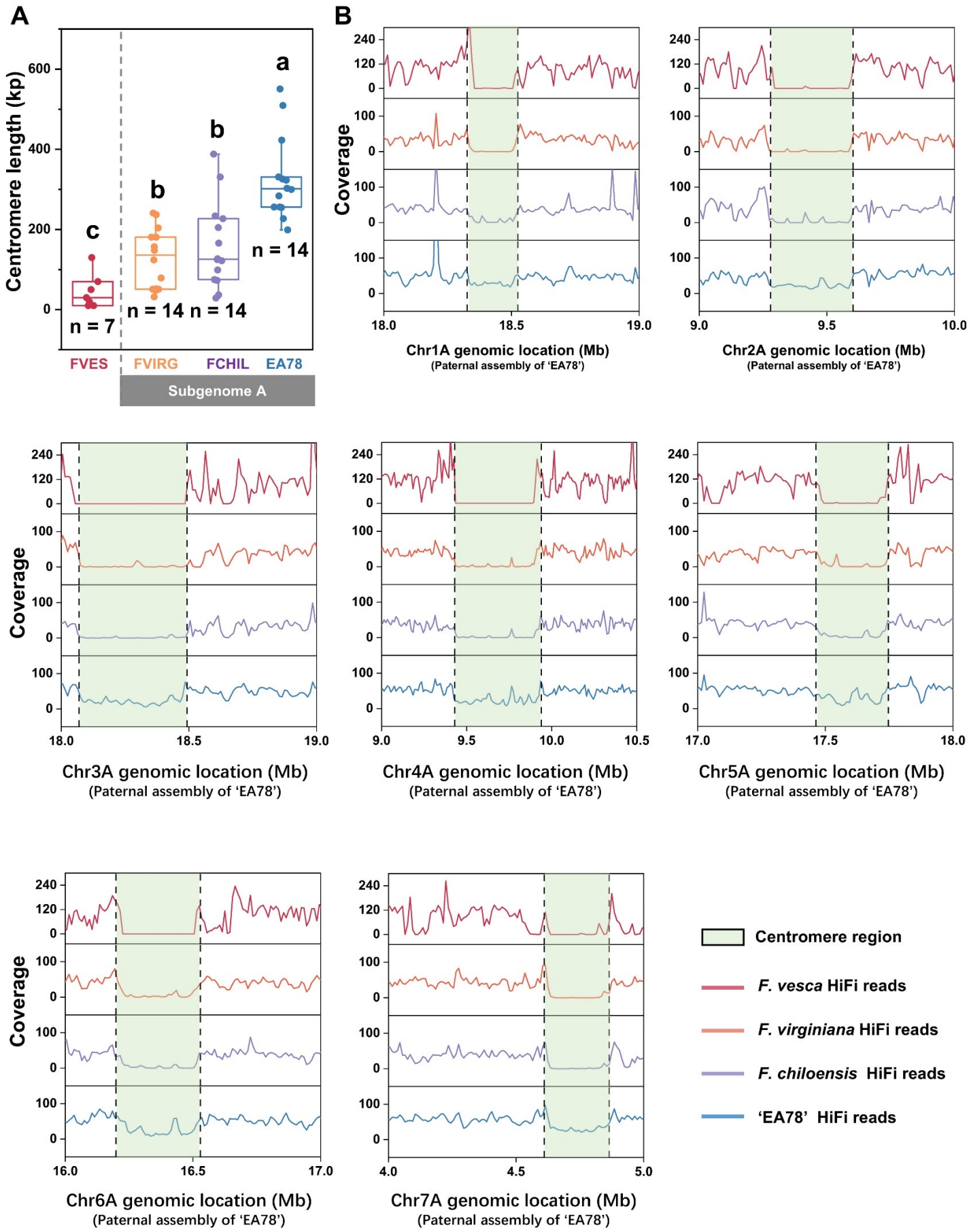


Fig. S27 | Comparison of the number of centromeric satellite counts in *F. vesca* and the A subgenome of octoploid strawberries. Each data point represents one chromosome. For some assemblies, one haplotype is available (n=7), while for others, both haplotypes are present (n=14). Note: Different colored lines represent different assemblies, with the red line representing the average connection.

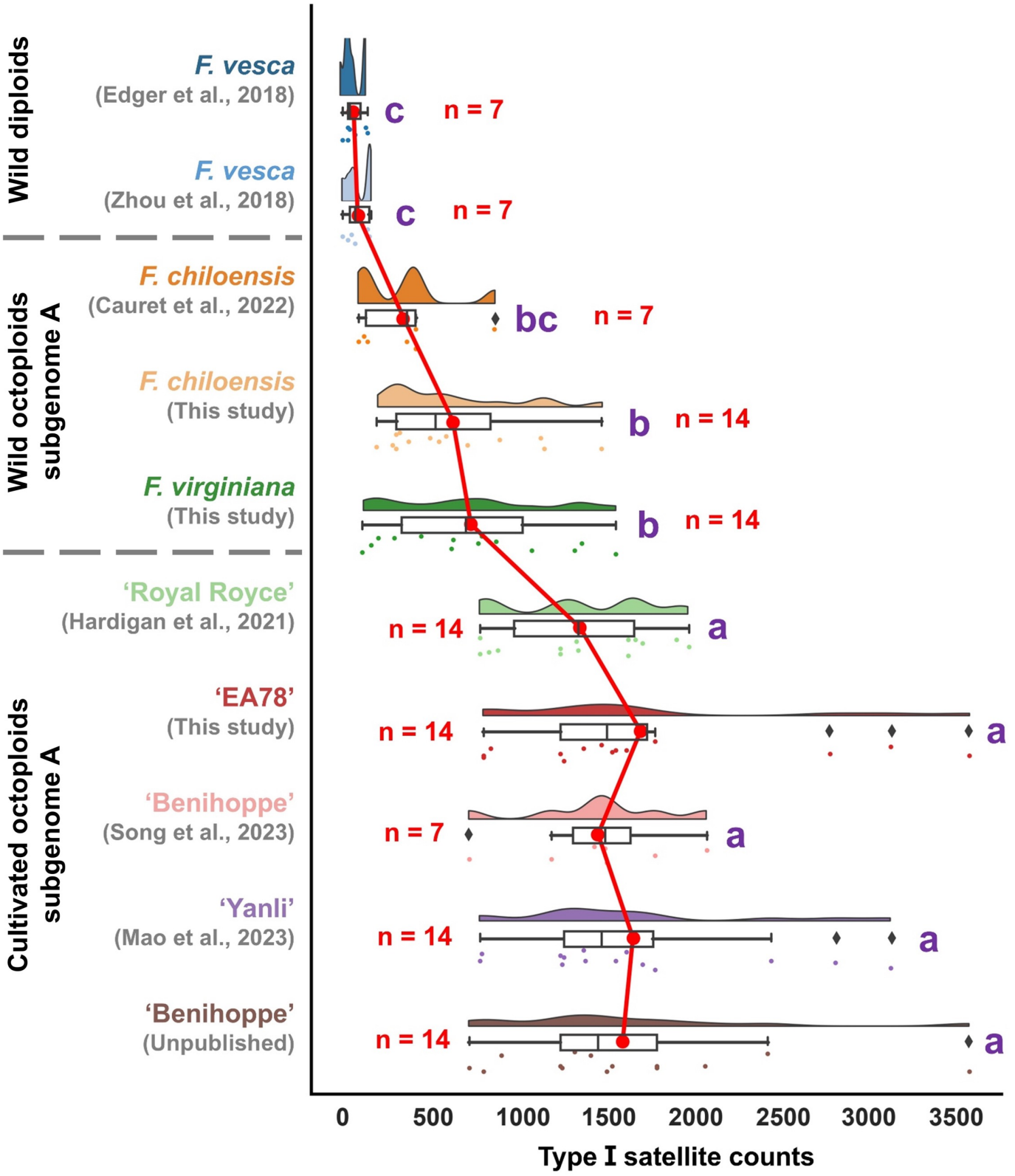


Fig. S28 | Identical centromeric bin analysis. (A) Distribution of centromeric bin similarity (Bin size: 150 bp). **(B)** Comparison of the percentage of identical centromeric bins among *F. vesca*, *F. chiloensis*, *F. virginiana* and 'EA78'. Note: Different colored lines represent different species: red for*F. vesca*, orange for *F. chiloensis*, purple for *F. virginiana,* and blue for 'EA78'.