
Supplementary information

**Chromosomal fusions trigger
rediploidization of autopolyploid genomes**

In the format provided by the
authors and unedited

Supplementary information

for “Chromosomal fusions trigger rediploidization of autopolyploid genomes”

Chuanshuai Xie^{1,#}, Zitu Ma^{1,#}, Chaowei Zhou^{1,#}, Kexin Ma^{1,#}, Haoyu Wang¹, Jiahong Wu¹, Yan Zhou¹, Yongrui Lu¹, Da Ji¹, Xuedie Gu², He Gao¹, Juntao Li¹, Suxing Fu¹, Weiqiang Li¹, Zhaofang Han¹, Shijun Xiao³, Fei Liu⁴, Benhe Zeng⁴, Shen’ao Chen⁵, Jiangong Niu⁶, Tao Zhang⁶, Jian Shen⁶, Chunna Liu⁷, Jing Luo², Daniel J. Macqueen⁸, Axel Meyer^{9,10,11,*}, Haiping Liu^{1,*}, Luohao Xu^{1,*}

*Correspondence:

luohaox@swu.edu.cn, luihappy@163.com, axel.meyer@uni-konstanz.de

equal contribution

Contents

Supplementary Figure 1 Morphological characteristics and altitude distribution of three group of snow carps.

Supplementary Figure 2 Kmer-distribution of hexaploid snow carp (*Schizothorax wangchiachii*).

Supplementary Figure 3 *S. younghusbandi* trio construction and assembly.

Supplementary Figure 4 Comparison of the read-binning assembly and the trio-based assembly.

Supplementary Figure 5 Landscape of transposable element (TEs) for *S. younghusbandi*.

Supplementary Figure 6 Selection pressure analysis of fused and non-fused quartets.

Supplementary Figure 7 The distribution of alternative allele depth for tetrasomic and disomic quartets.

Supplementary Figure 8 Distribution of bivalent and multivalent inheritance genotypes across chromosomes.

Supplementary Figure 9 Divergence time estimates for wave2 based on an independent rate clock model.

Supplementary Figure 10 The karyotype of meiotic metaphase I cells in the testis.

Supplementary Figure 11 Features of *S. younghusbandi* centromere.

Supplementary Figure 12 Average sequence divergence (*Ks*) among all pairwise homologous in short and long arm.

Supplementary Figure 13 Gene loss in *S. younghusbandi*.

Supplementary Figure 14 Expression patterns of *uf* and *f*ohnologs in eleven tissues.

Supplementary Figure 15 Gene expression level between ohnolog *uf* and *f*.

Supplementary Figure 16 Phylogenetic tree topologies on chromosome 19 and 22.

a

non-specialized group
also 'basal' group
(*Schizothorax* and *Aspidorhynchus*)



Two pairs of barbels
The body is covered with small scales
Three rows of pharyngeal teeth

specialized group
(*Diptychus*, *Gymnodiptychus*, and *Ptychobarbus*)



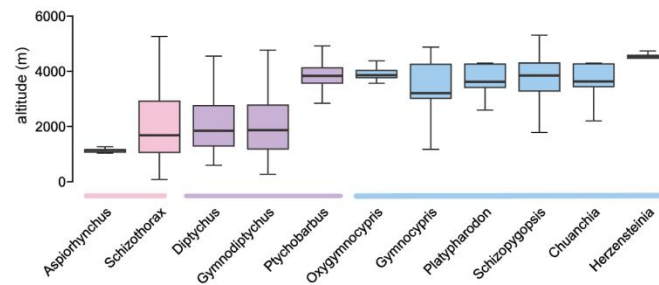
One pair of barbels
Some of the scales are degenerated
Two rows of pharyngeal teeth

highly specialized group
(*Oxygymnocypris*, *Gymnocypris*, *Schizopygopsis*,
Chuanchia, *Platypharodon* and *Herzensteinia*)

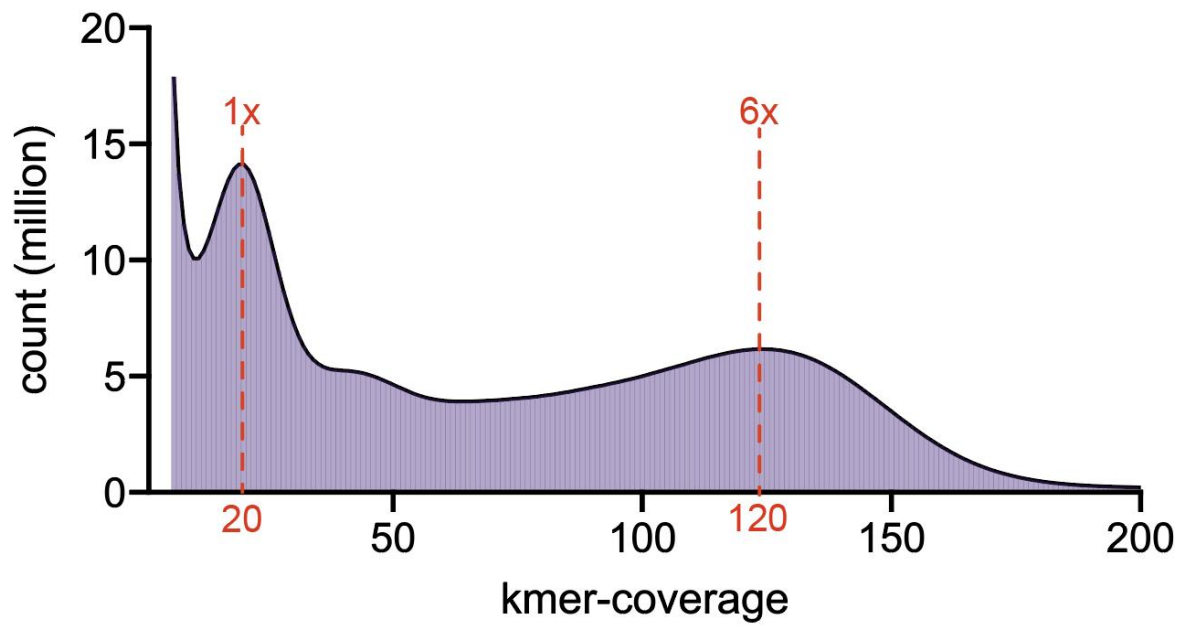


No barbels
No scales
One to two rows of pharyngeal teeth

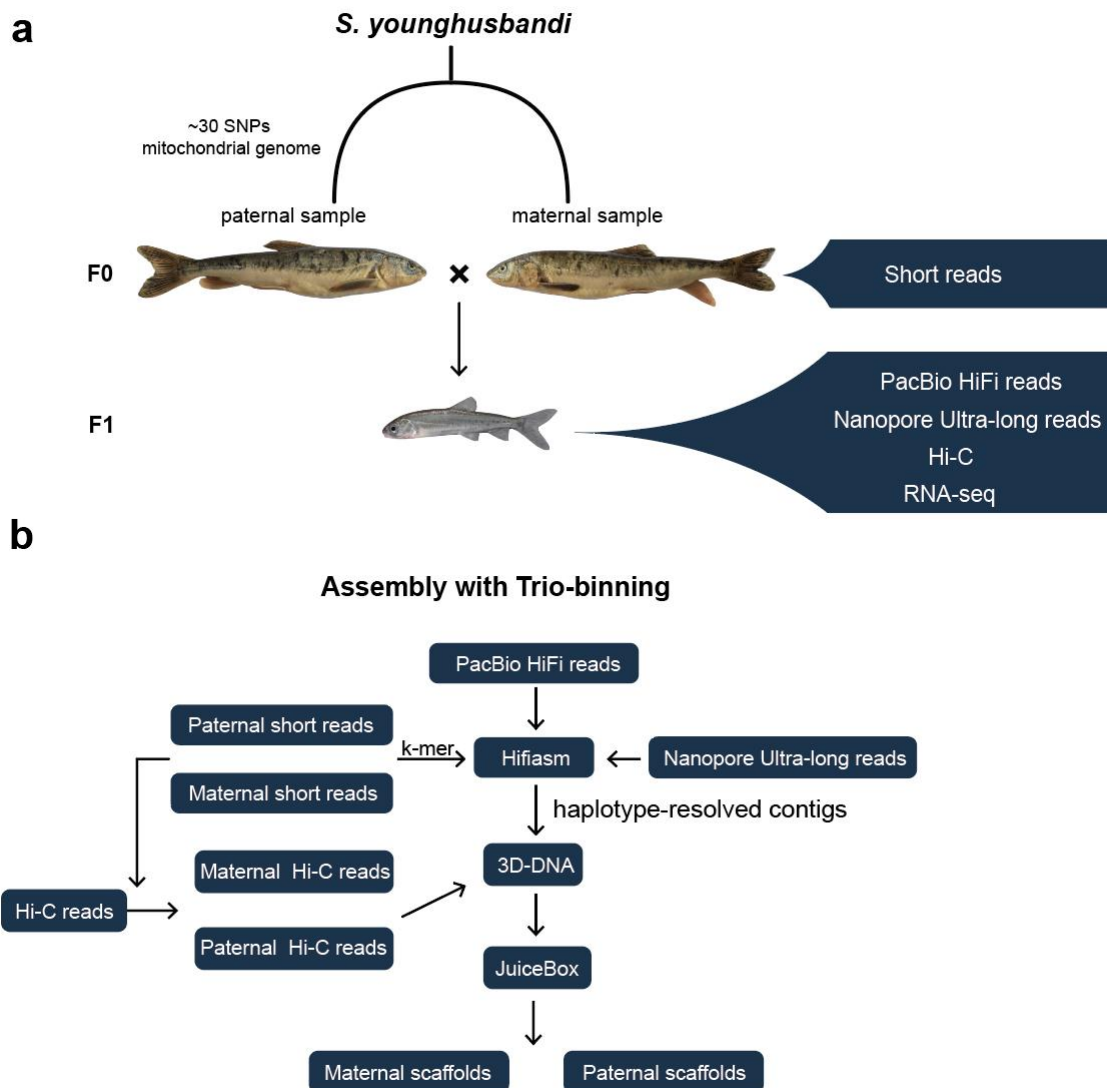
b



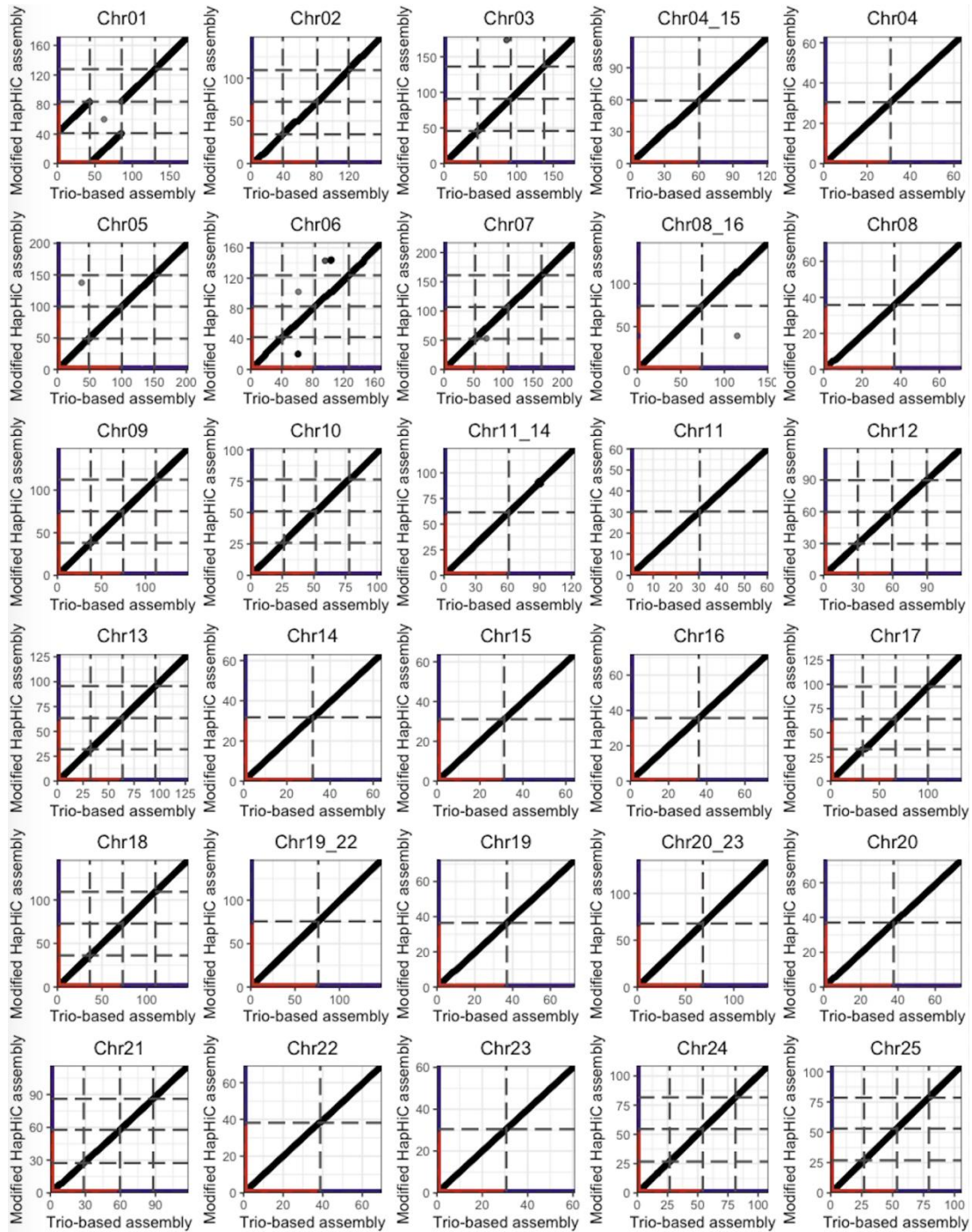
Supplementary Figure 1 Morphological characteristics and altitude distribution of three group of snow carps. (a) Morphological features of three species of schizothoracine fish: the left side shows the whole body, and the right side shows a close-up of the head (photo credit, He Gao). (b) Altitude distribution of 11 genera of snow carps. The color represent different morphological groups.



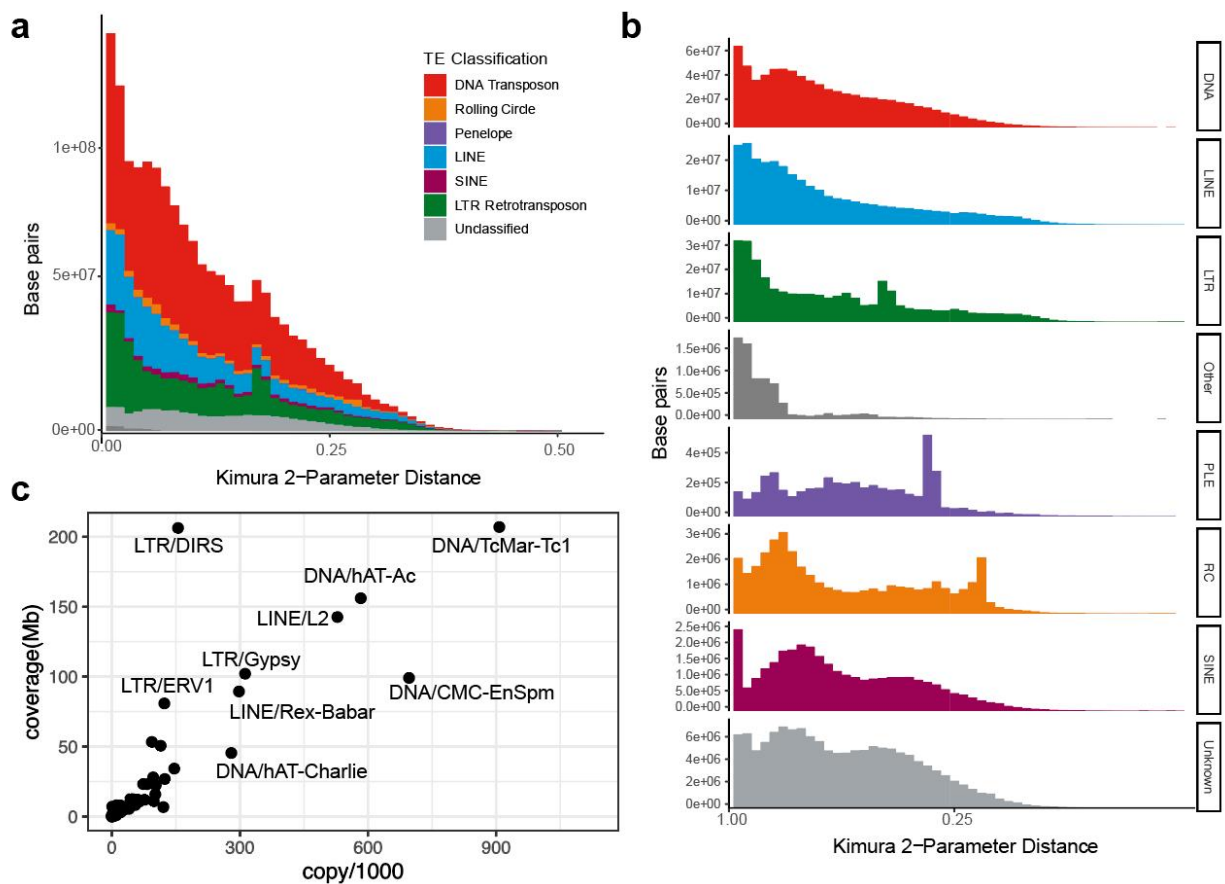
Supplementary Figure 2 Kmer-distribution of hexaploid snow carp (*Schizothorax wangchiachii*).



Supplementary Figure 3 *S. younghusbandi* trio construction and assembly. (a) The trio of *S. younghusbandi* consists of two parents with mitochondrial genomes differing by 30 SNPs and their offspring (photo credit, He Gao). (b) Workflow of assembly with trio-binning.

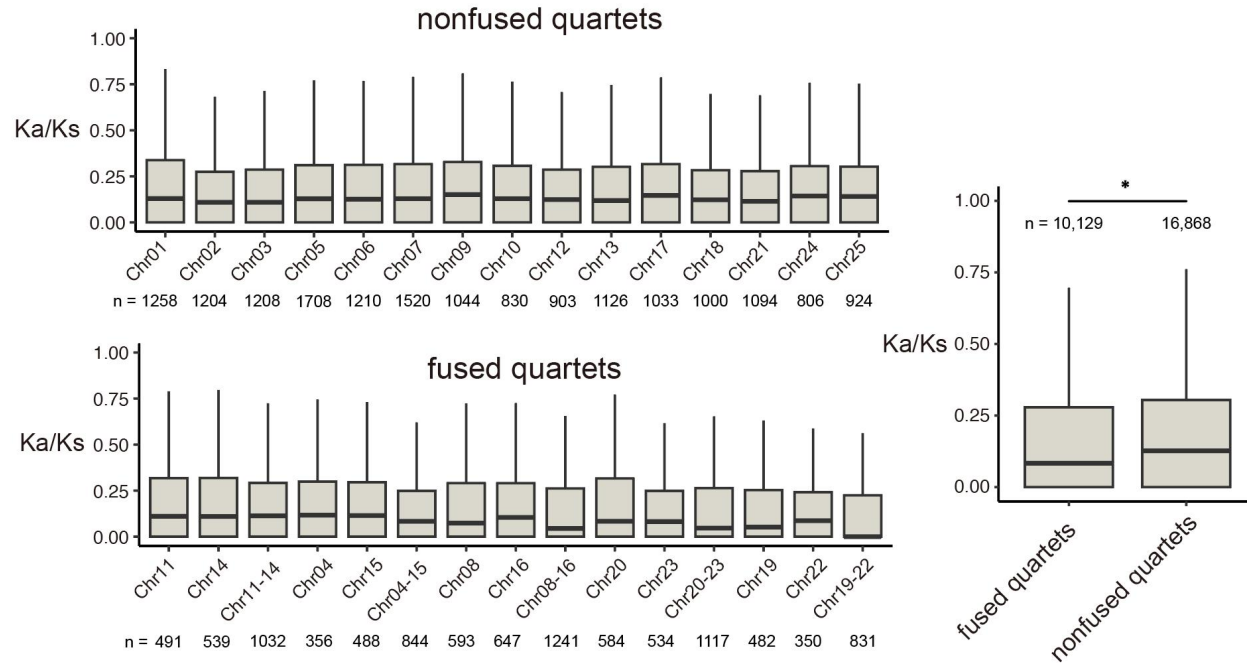


Supplementary Figure 4 Comparison of the read-binning assembly and the trio-based assembly. The dot plot shows the collinearity of the chromosomes from two assembly version. The x-axis is trio-based assembly, and the y-axis is read-binning assembly. Red and purple represent haplotypes from maternal and paternal, respectively.

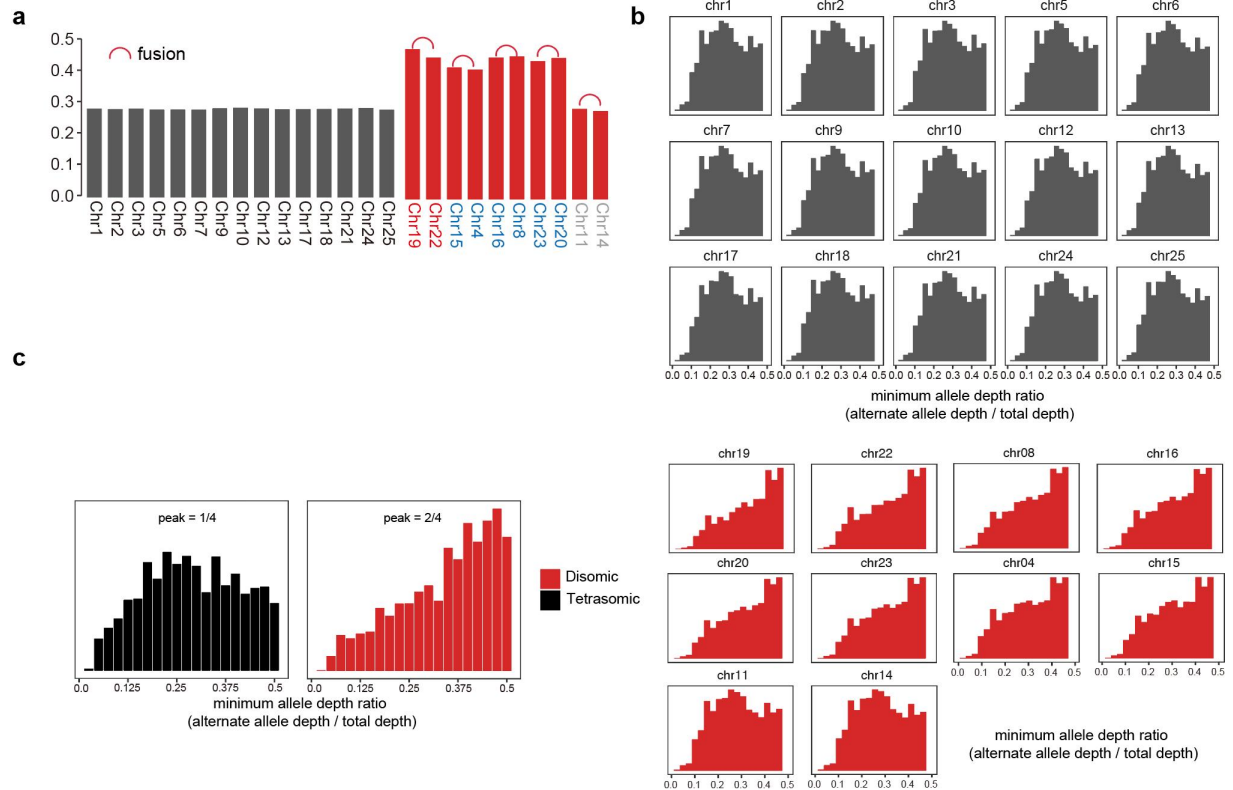


Supplementary Figure 5 Landscape of transposable element (TEs) for *S. younghusbandi*.

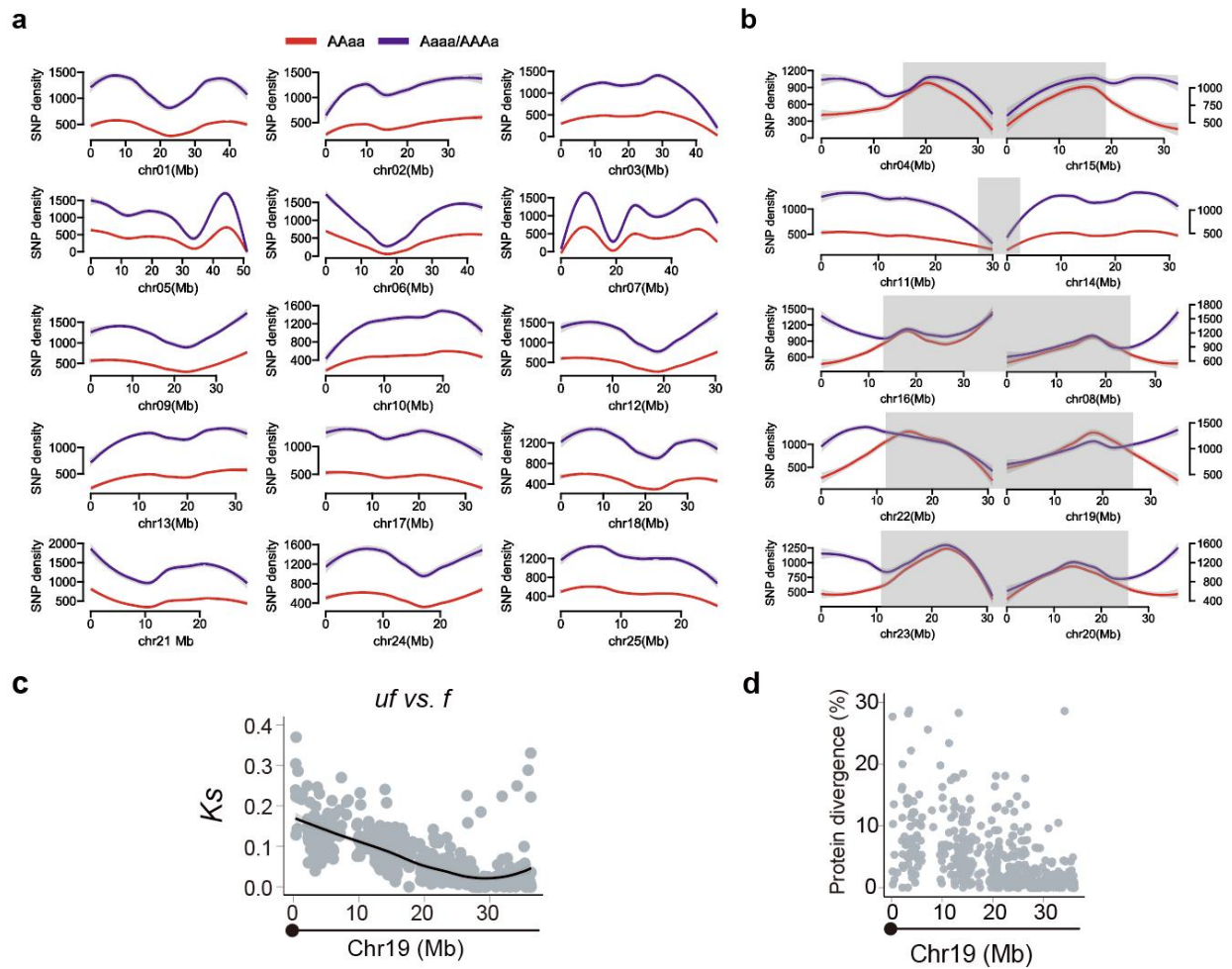
(a) and (b) Distribution of kimura distance of each type of TEs. Different colors represent different TE classe. (c) Scatter plot of copy number (x-axis) versus total length in base pairs (y-axis) for individual TE families. Abundant families are highlighted.



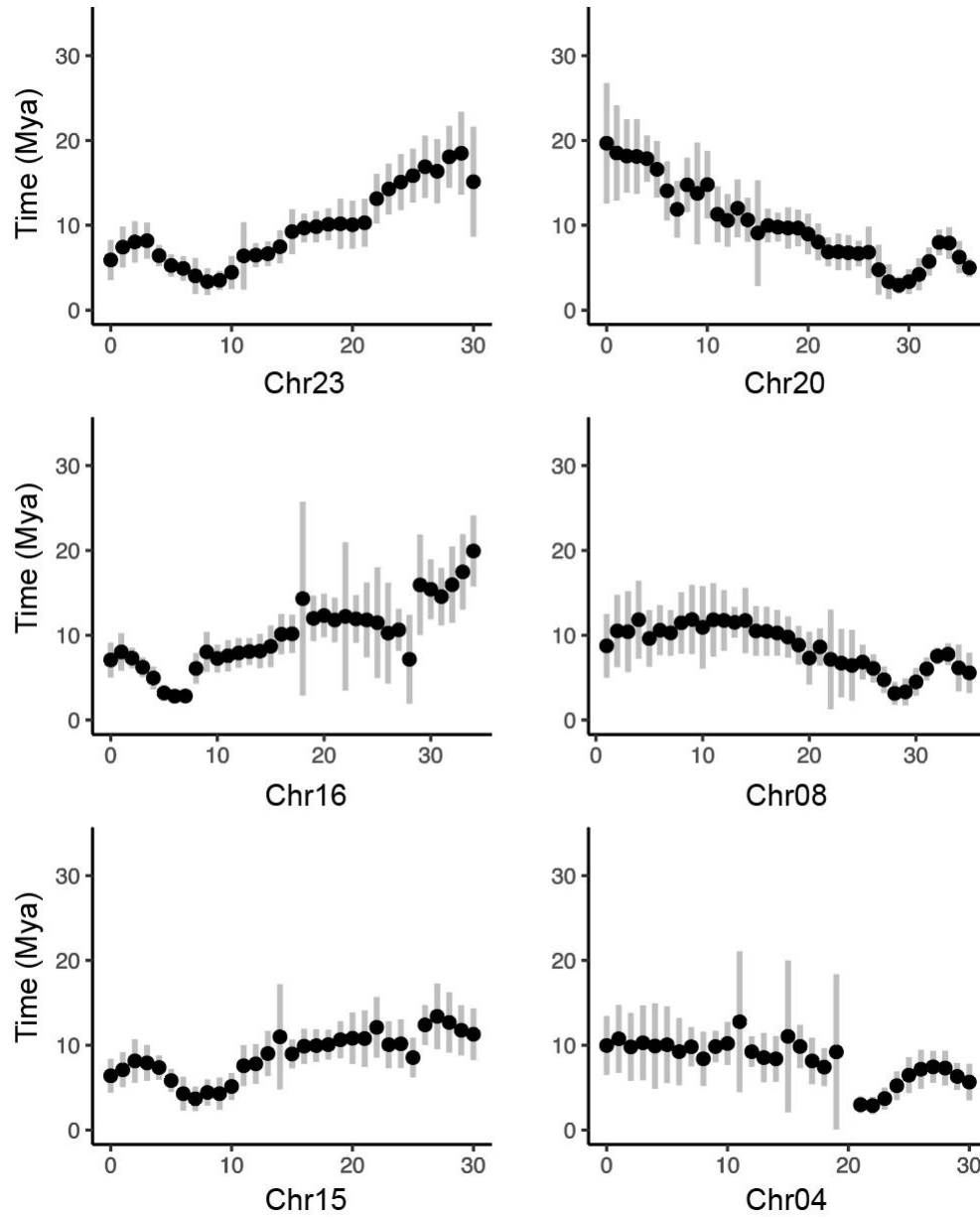
Supplementary Figure 6 Selection pressure analysis of fused and non-fused quartets. The left panel shows the Ka/Ks ratio for each chromosome, while the right panel presents the aggregated distribution. Asterisks denote statistical significance (two-sided Mann-Whitney U test, $p < 0.01$). The chromosome name Chr11-14, Chr04-15, Chr08-16, Chr20-23 and Chr19-22 indicate the fused chromosomes (f copy). Each dot represents one gene pair (n indicated above/below). The box plots display the median (center line), the interquartile range (box), and the data range within $1.5 \times IQR$ (whiskers). The sample size (n) denotes the number of homologous gene pairs used in the calculation, as labeled below.



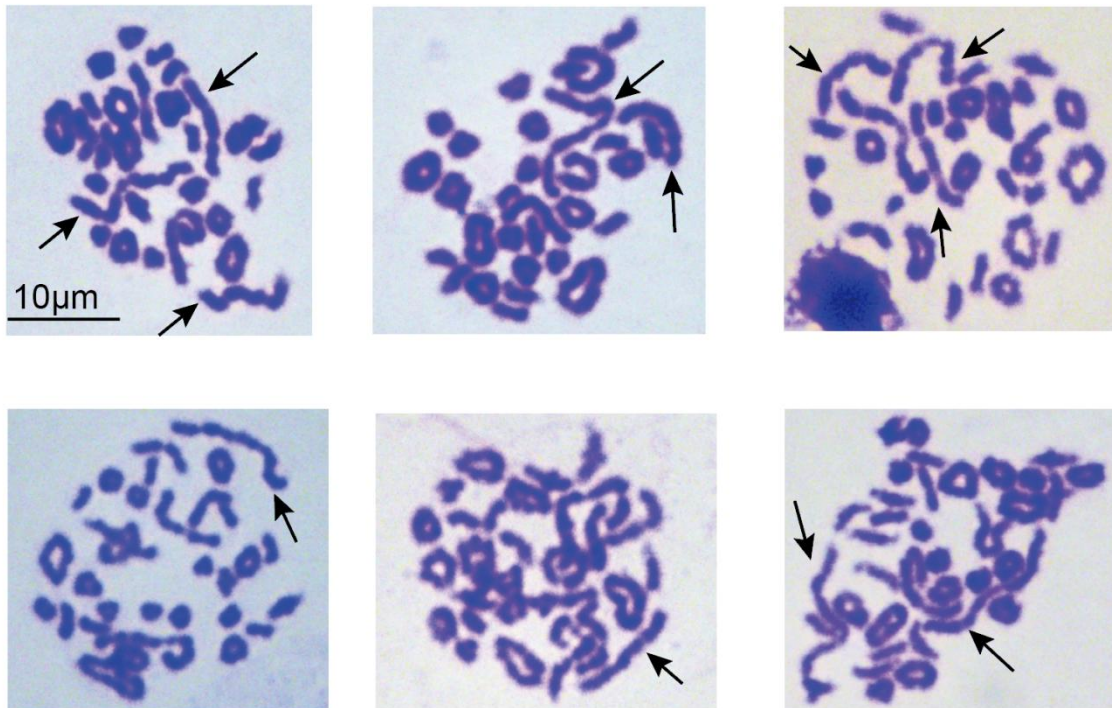
Supplementary Figure 7 The distribution of alternative allele depth for tetrasomic and disomic quartets. (a) Ratio of disomic genotype (AAaa) among all heterozygous tetraploid genotype. The red line represents the 5 pairs of fused chromosomes. The three waves of rediploidization are indicated by the colors of the chromosome names. Two-sided Mann-Whitney U test was used to compare the differences between fused and unfused quartets. (b) Alternative allele depth distribution for non-fused quartets (top) and fused quartets (bottom). (c) The distinct distribution pattern for tetrasomic and disomic quartets shows that the peak of the depth ratio is close to 0.25 under tetrasomic inheritance, whereas it is close to 0.5 under disomic inheritance. Among them, Chr11 and Chr14, despite being fused quartets, have not yet exhibited a disomic inheritance pattern, as they may have undergone recent fusion. All tetrasomic chromosomes are highlighted in red.



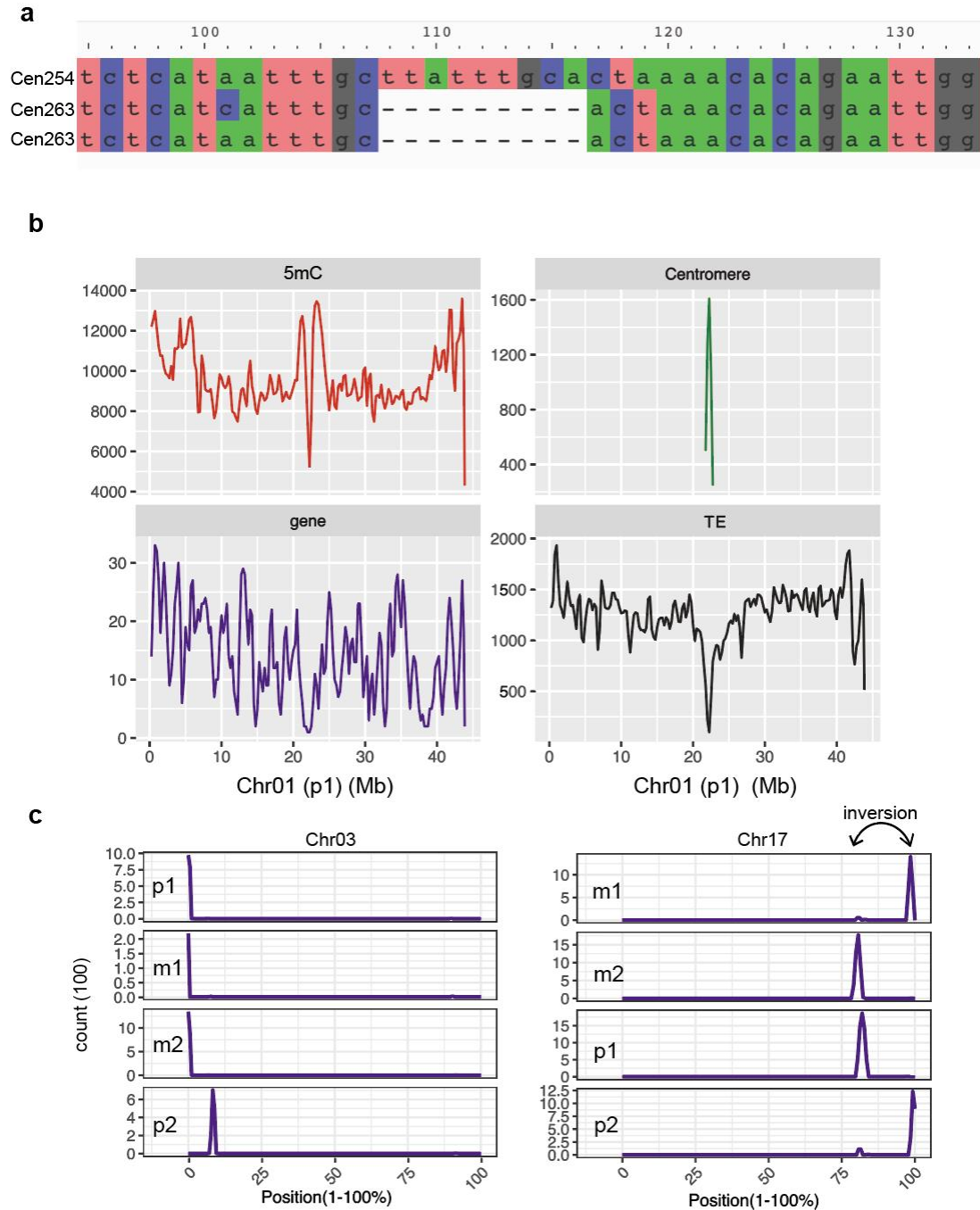
Supplementary Figure 8 Distribution of bivalent and multivalent inheritance genotypes across chromosomes. (a) non-fused quartets (b) fused quartets. The tetrasomic and disomic genotypes are labeled in blue and red respectively. (c) The distribution of K_s values between homologs (unfused vs. fused) along the chromosome is elevated at the fusion site. (d) The distribution of protein sequence divergence is elevated at the fusion site. The black dots indicate the locations of fusion sites.



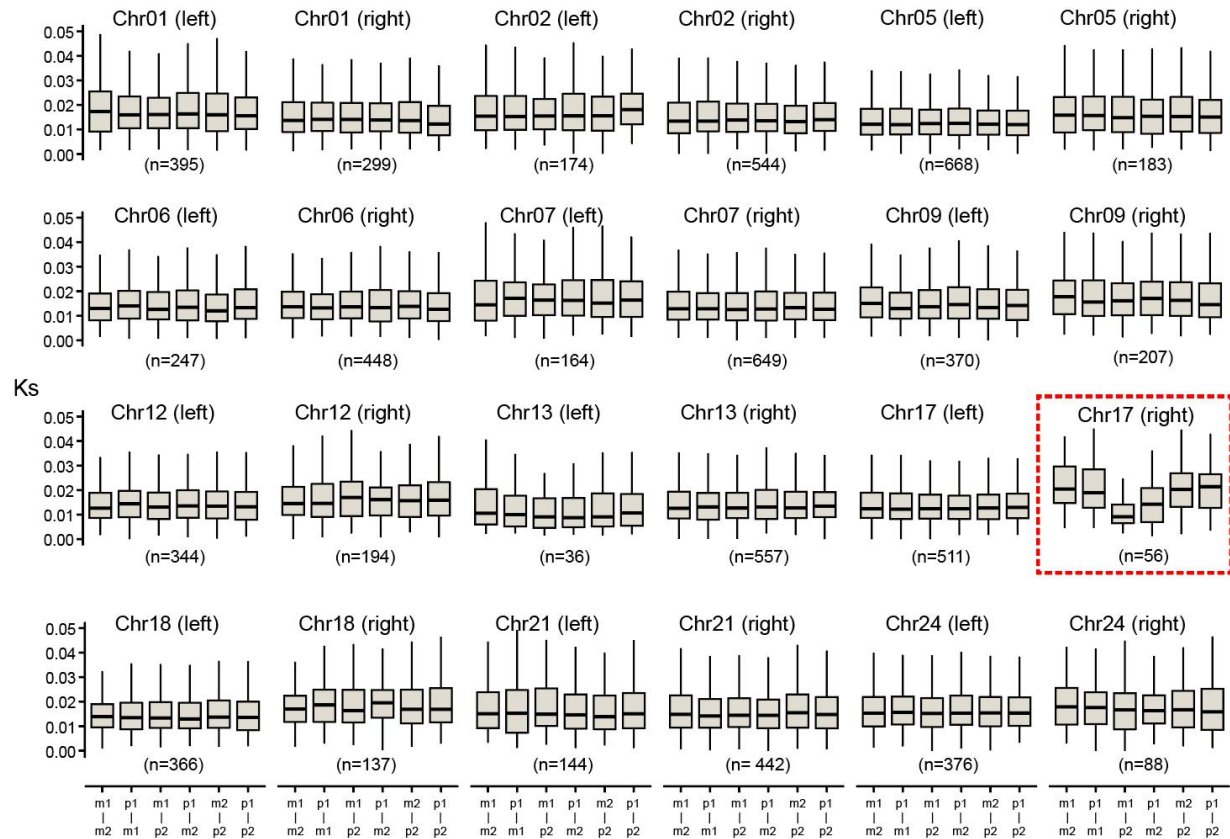
Supplementary Figure 9 Divergence time estimates for wave2 based on an independent rate clock model. For each genomic window, the point shows the median estimate of node age, and the bar indicates the 95% HPD interval.



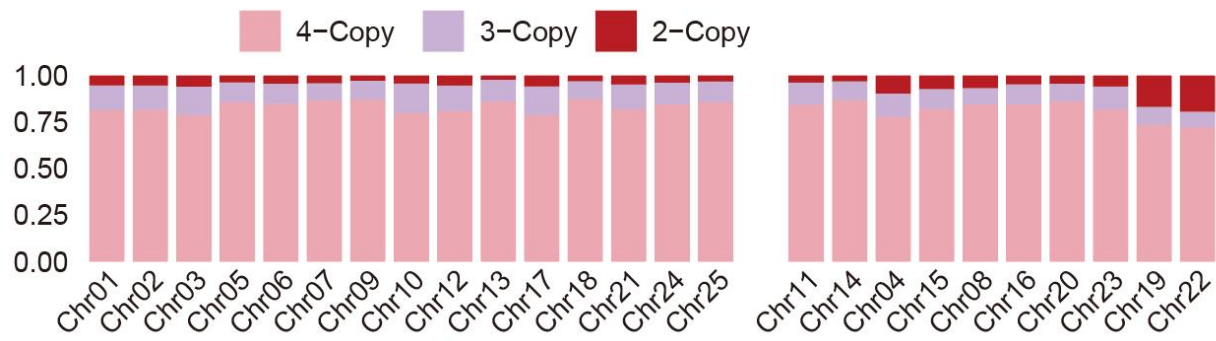
Supplementary Figure 10 The karyotype of meiotic metaphase I cells in the testis. The arrows point to the chain-shaped pairing structures. Multiple centromere constrictions can be observed in those long chain pairing structures.



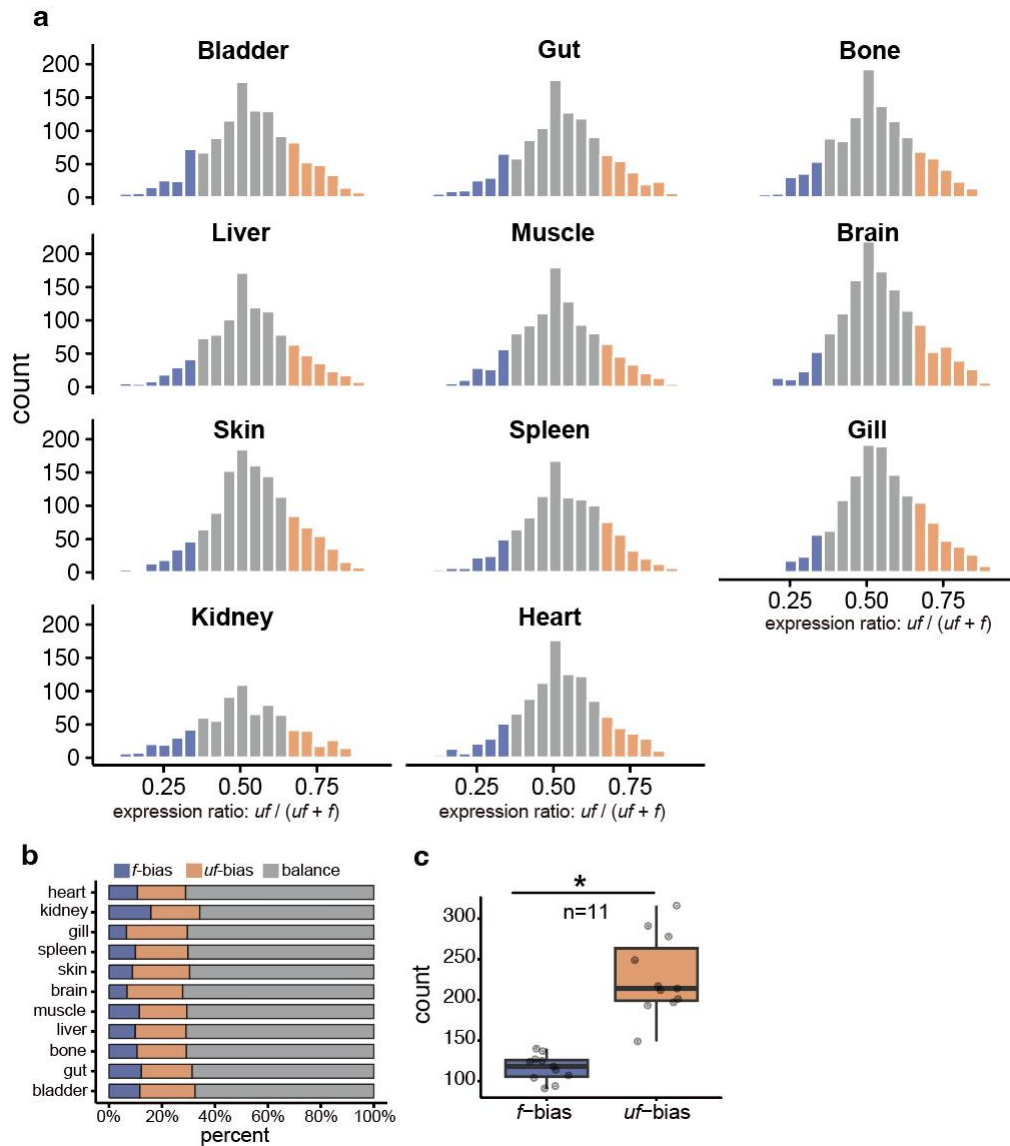
Supplementary Figure 11 Features of *S. younghusbandi* centromere. (a) Sequence differences between Cen254 and Cen263. Different colored backgrounds represent different bases (visualized using AliView). (b) Distribution of DNA methylation (red line), centromeric repeats (green line), gene density (purple line), and transposable element density (black line) on Chr01 (p1). (c) Centromeric inversion on Chr03 and Chr17.



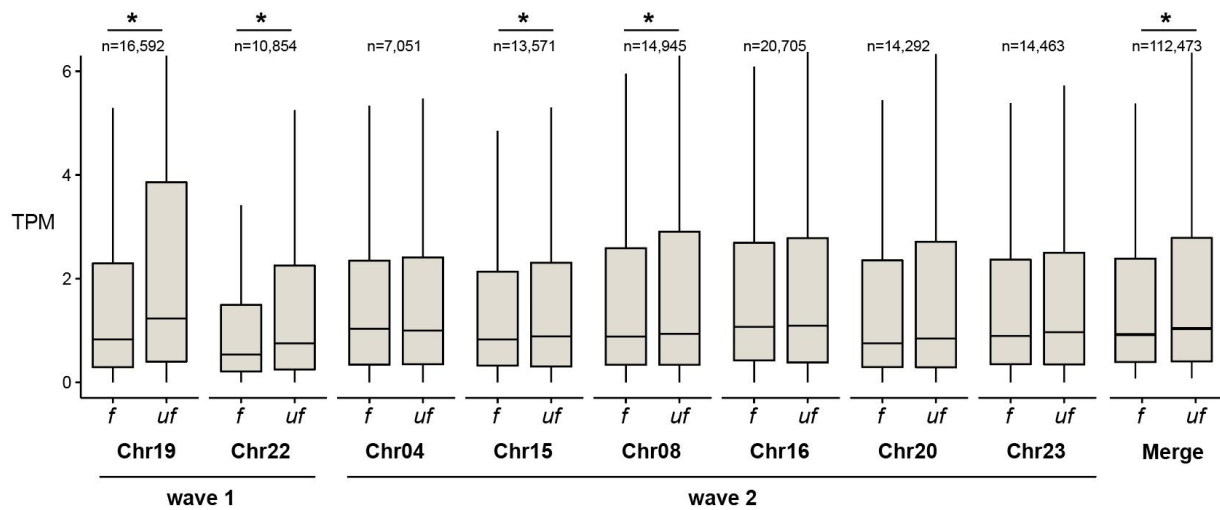
Supplementary Figure 12 Average sequence divergence (K_s) among all pairwise homologous in short and long arm. In the unfused quartets, except for Chr17 (marked in red), the long and short arms exhibited consistent patterns. "Left" and "right" denote the left- and right-flanking regions of the centromere, corresponding to the long or short arm. Centromere positions are shown in Extended Data Fig. 3d. The three telocentric chromosomes (Chr03, Chr10, and Chr25) were excluded from the analysis. The box plots display the median (center line), the interquartile range (box), and the data range within $1.5 \times \text{IQR}$ (whiskers). The sample size (n) denotes the number of homologous gene pairs used in the calculation, as labeled below.



Supplementary Figure 13 Gene loss in *S. youngusbandi*. The proportion of 2-copy, 3-copy, and 4-copy genes between tetrasomic chromosomes and disomic chromosomes. The left chromosomes is tetrasomic and the right chromosomes is disomic.



Supplementary Figure 14 Expression patterns of *uf* and *f* ohnologs in eleven tissues. (a) Distribution of the expression ratios of *uf* ohnolog. Ohnologs with the *uf* copy contributing more than 0.65 of the total expression are defined as *uf*-bias (marked in orange), those with ratio lower than 0.35 are defined as *f*-bias (marked in blue); and the remaining ohnologs are classified as balanced (marked in grey). (b) Proportion of three ohnolog expression patterns across eleven tissues. (c) Boxplot shows that the proportion of the *uf*-biased pattern is significantly higher than that of the *f*-biased pattern (two-tailed paired t-test, p -value < 0.05). Plots display the median (center line), interquartile range (box), and the data range within 1.5×IQR (whiskers). The sample size (n) denotes the number of pairwise comparisons across 11 tissues.



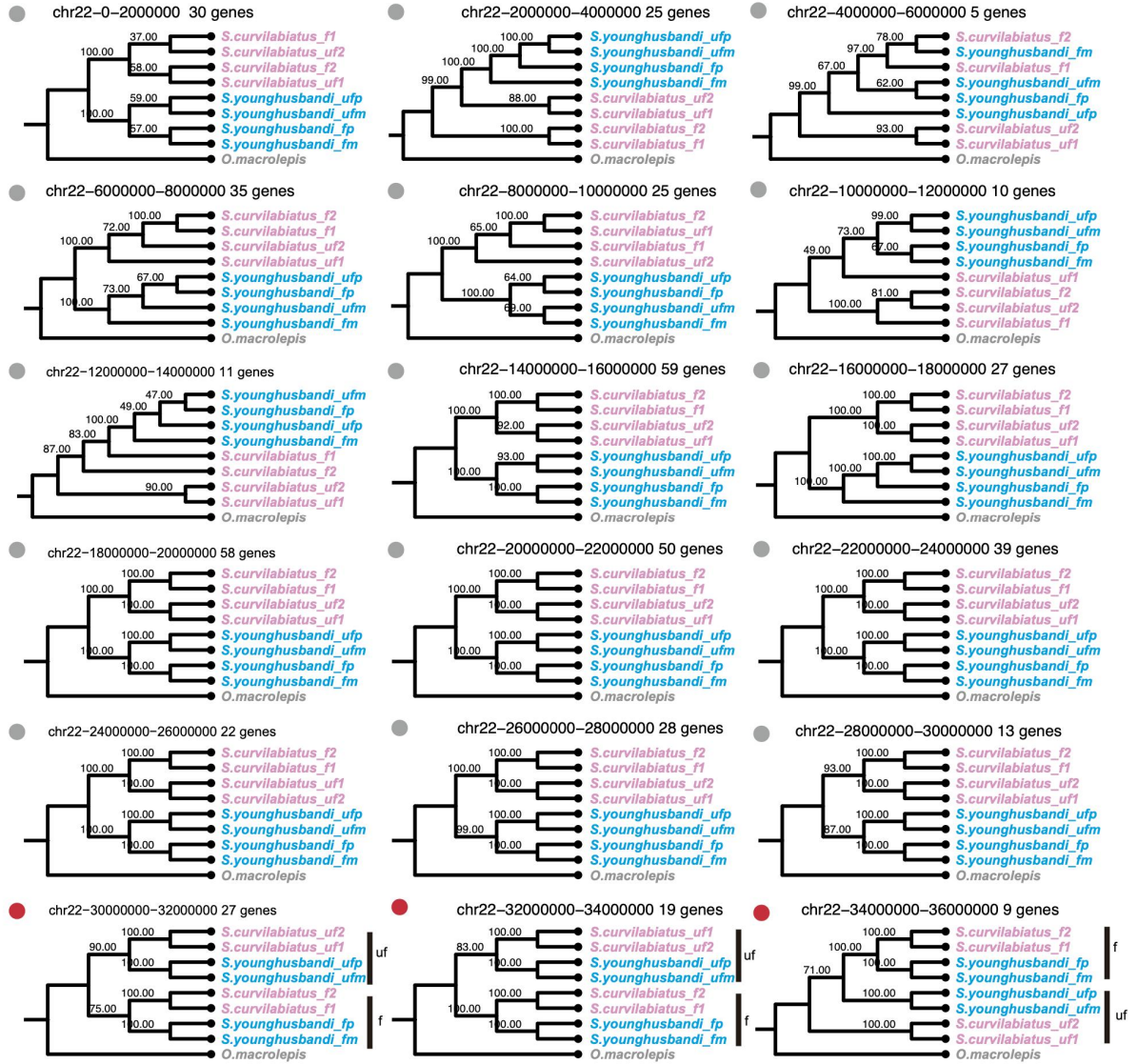
Supplementary Figure 15 Gene expression level between ohnolog *uf* and *f*. RNA-seq data across eleven tissues were combined and subjected to a two-tailed t-test, with asterisks indicating statistical significance at $p\text{-value} < 0.05$. The box plots display the median (center line), the interquartile range (box), and the data range within $1.5 \times \text{IQR}$ (whisker). The sample size (n) denotes the number of homologous gene pairs used in the calculation, as labeled above.

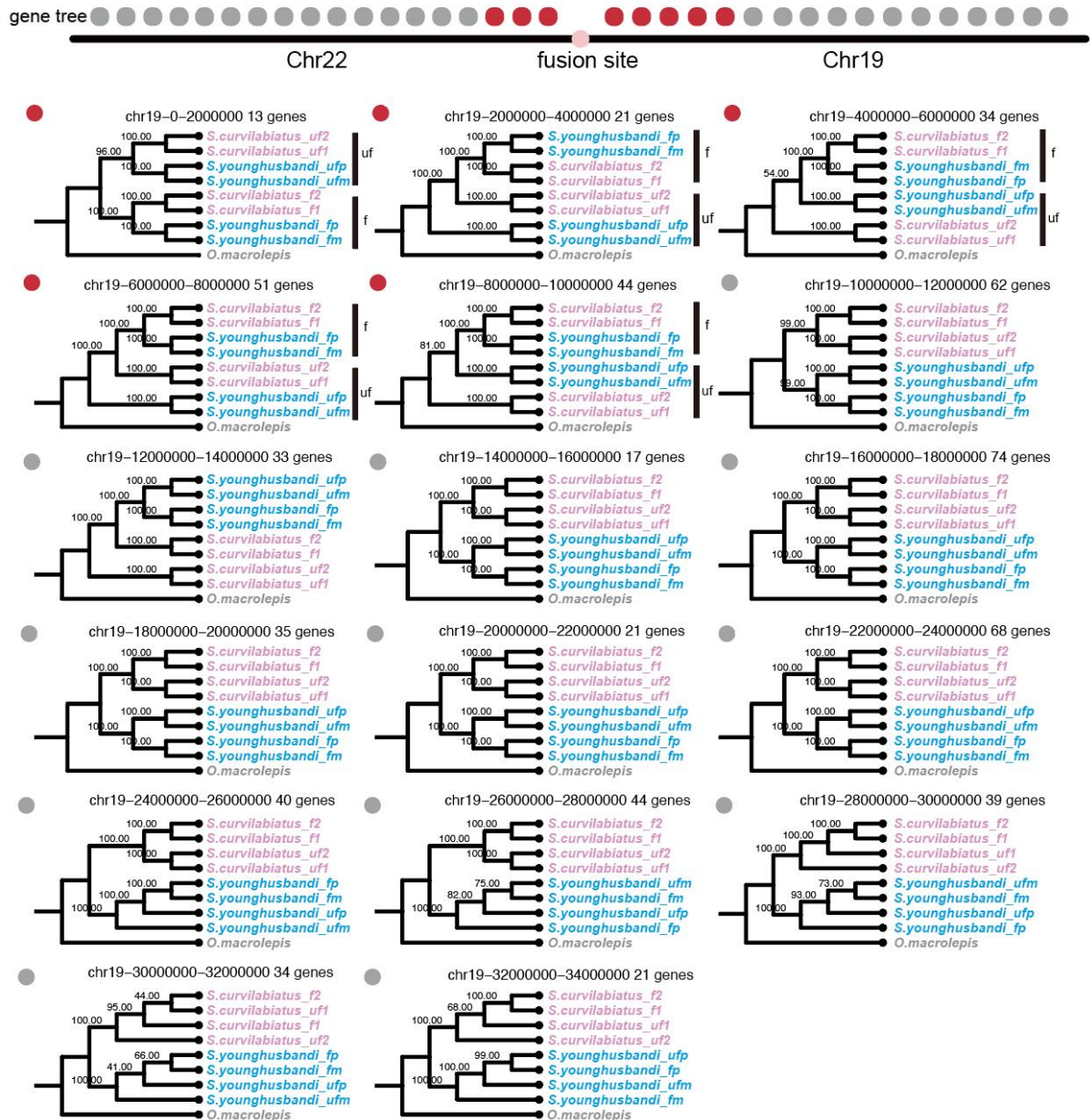


Chr22

fusion site

Chr19





Supplementary Figure 16 Phylogenetic tree topologies on chromosome 19 and 22. Maximum-likelihood phylogenetic trees were constructed for genes within each 2-Mb window using concatenated alignments (protein coding genes). The approximately unbiased (AU) test was applied to compare two alternative models: ancestral rediploidization and independent rediploidization. Tree topologies supporting ancestral rediploidization are highlighted in red, whereas those supporting independent rediploidization are shown in gray. Bootstrap support values are indicated on each tree, and branch colors denote different species. At the top, the positions of individual windows for Chr19 and Chr22 are shown, with points colored according to the model they support. The approximate locations of fusion sites are indicated by pink dots.