iScience, Volume *26*

Supplemental information

Pangenome obtained by long-read sequencing of 11 genomes reveal hidden functional structural variants in pigs

Yi-Fan Jiang, Sheng Wang, Chong-Long Wang, Ru-Hai Xu, Wen-Wen Wang, Yao Jiang, Ming-Shan Wang, Li Jiang, Li-He Dai, Jie-Ru Wang, Xiao-Hong Chu, Yong-Qing Zeng, Ling-Zhao Fang, Dong-Dong Wu, Qin Zhang, and Xiang-Dong Ding

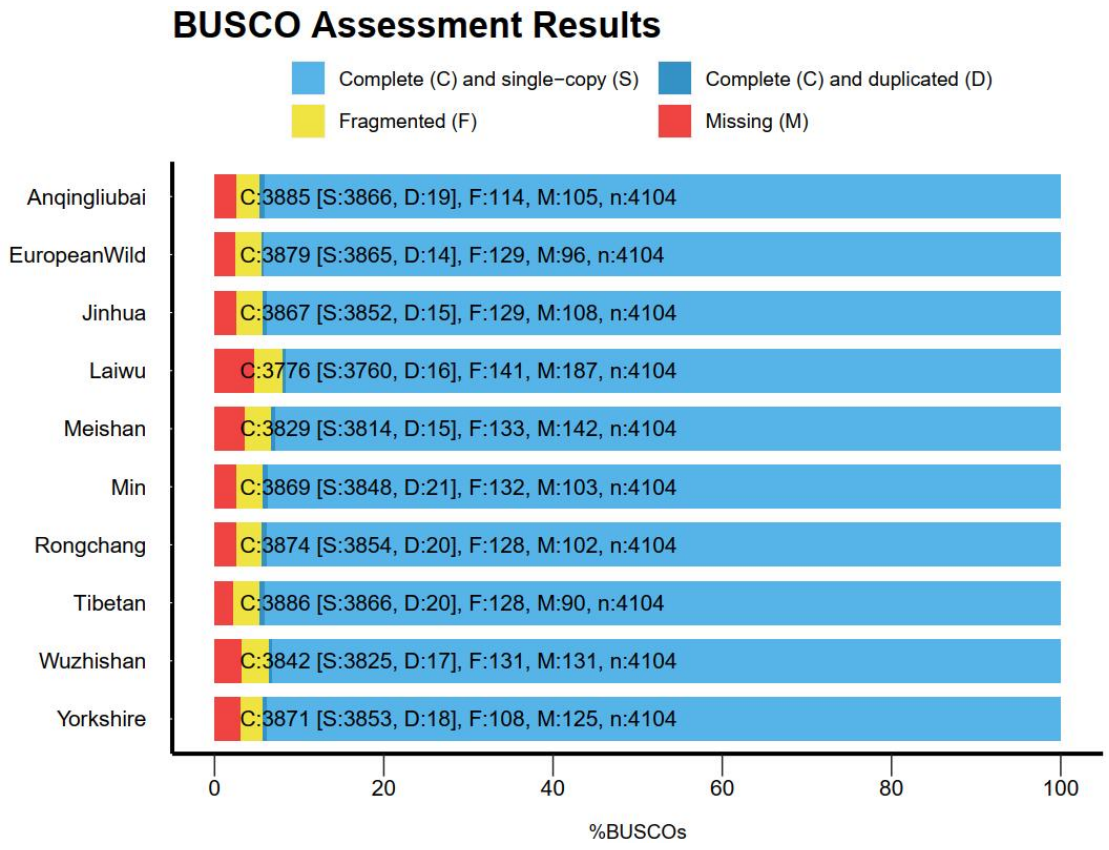

Figure S1. BUSCO assessments of the de novo assembly contigs. Related to Figure 1 and STAR Methods.

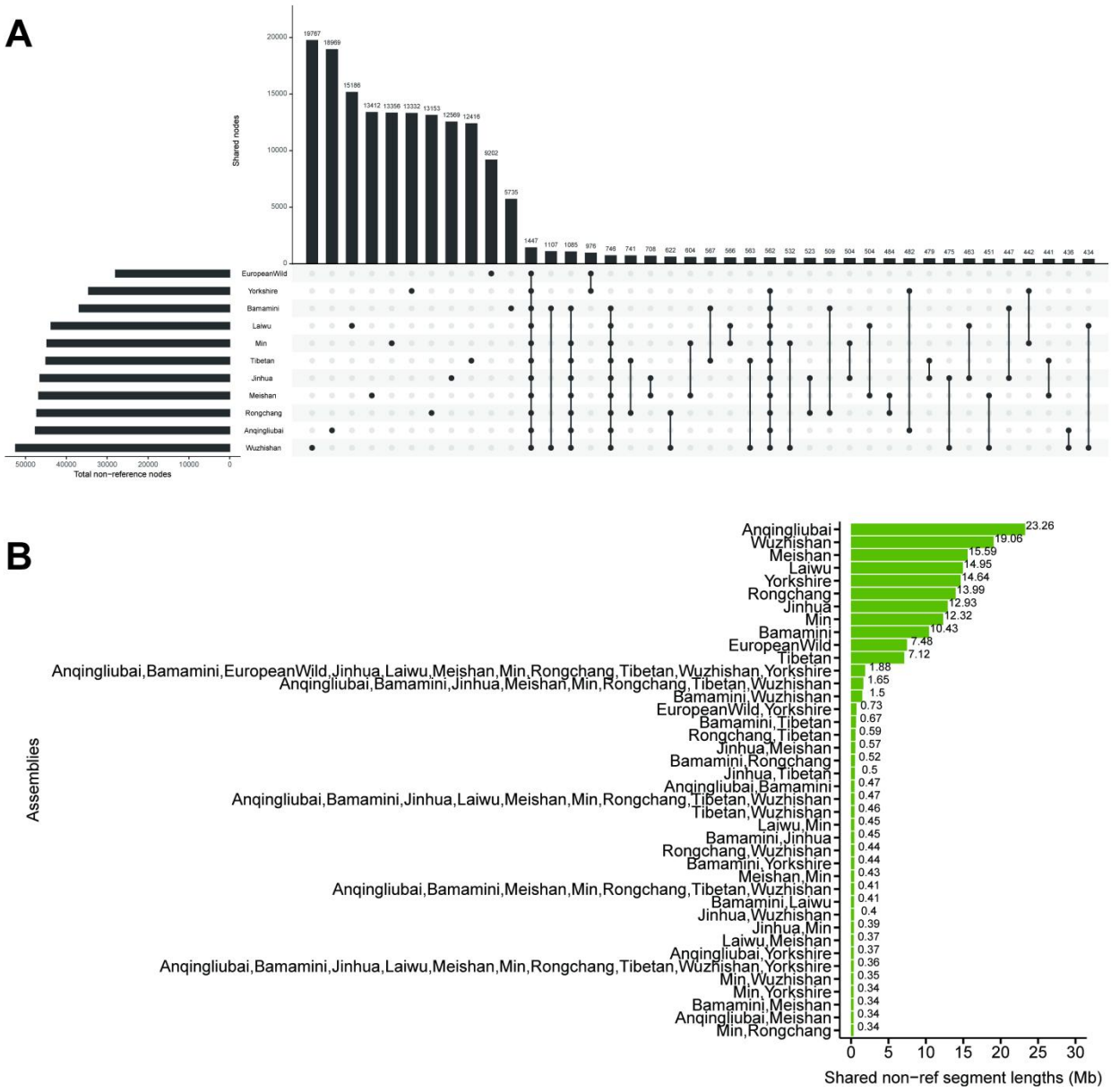

Figure S2. Non-reference nodes stats across assemblies. Related to Figure 1 and STAR Methods. Intersection of non-reference nodes (A) and the cumulative length of non-reference nodes (B) of the 11 assemblies. Only the top 40 are on display.

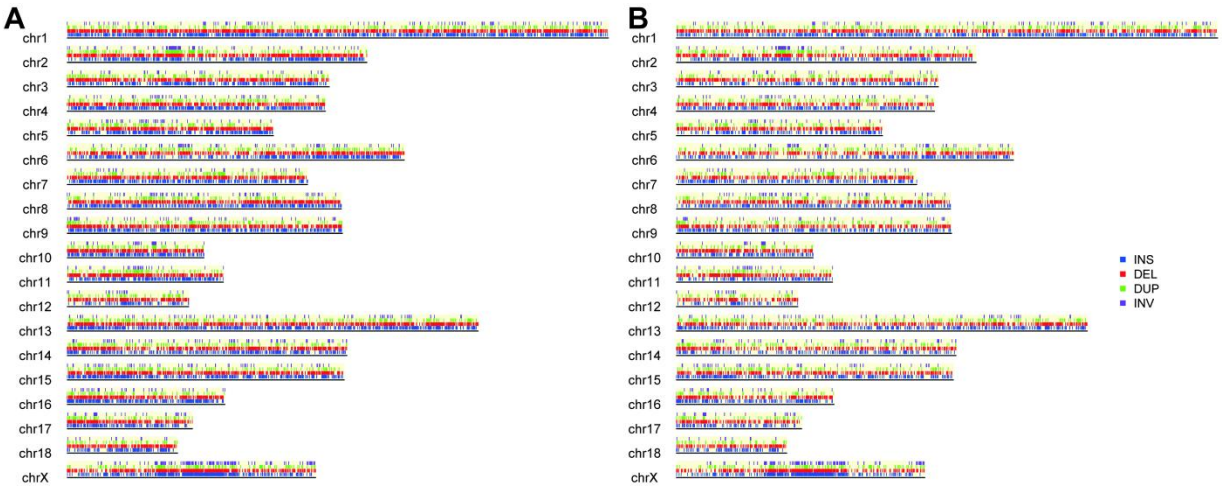

Figure S3. Genome-wide distribution of large SVs (>= 1 kb) among the pig chromosomes. Related to Figure 2 and STAR Methods. All the non-redundant SVs as shown in (A) and Chinese specific SVs as shown in (B).

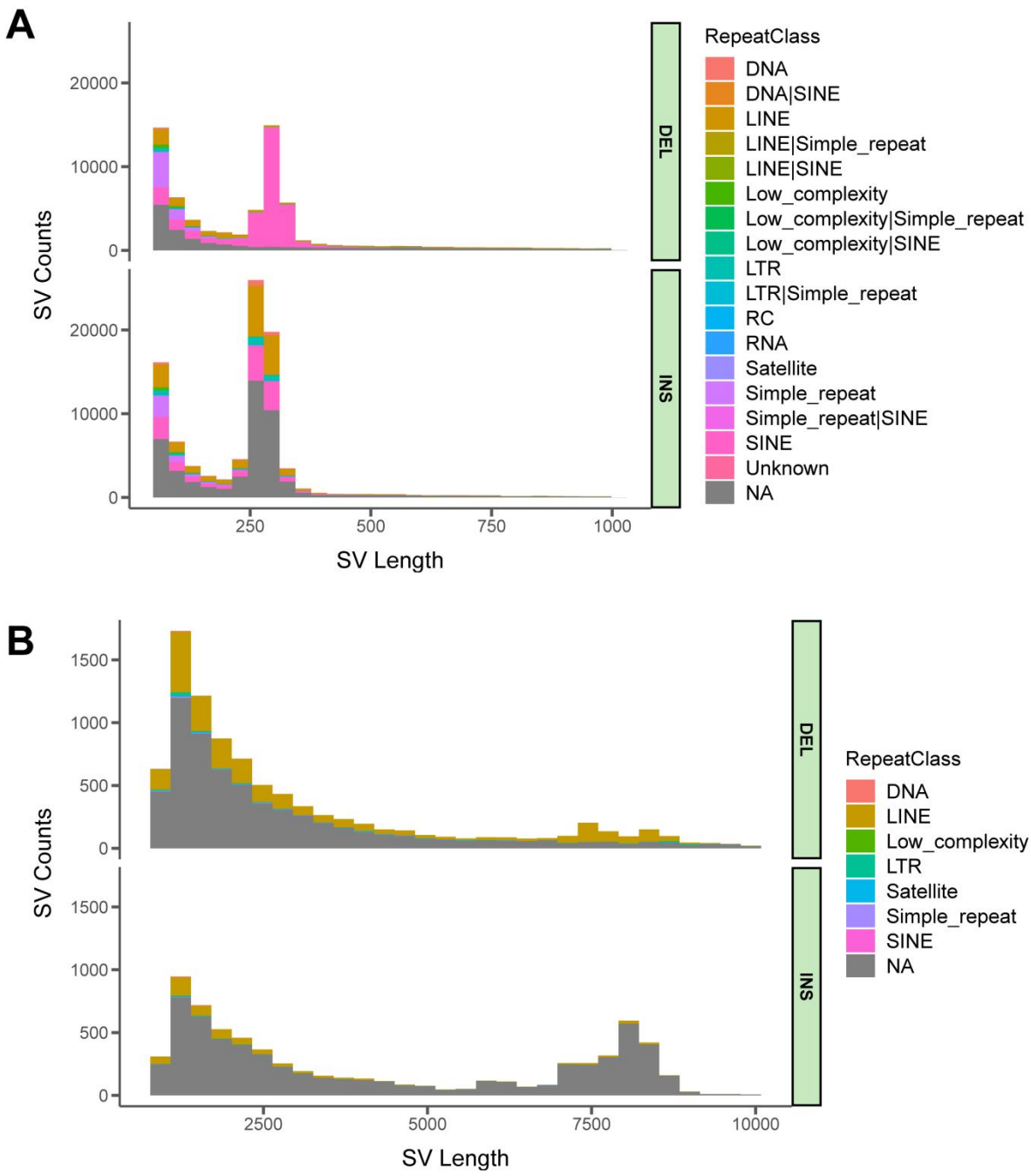

Figure S4. SV length distribution. Related to Figure 2 and STAR Methods. Length of INSs and DELs classified by intersected repeat elements across the ranges of 50 bp to 1 kb (A) and 1 kb to 10 kb (B).

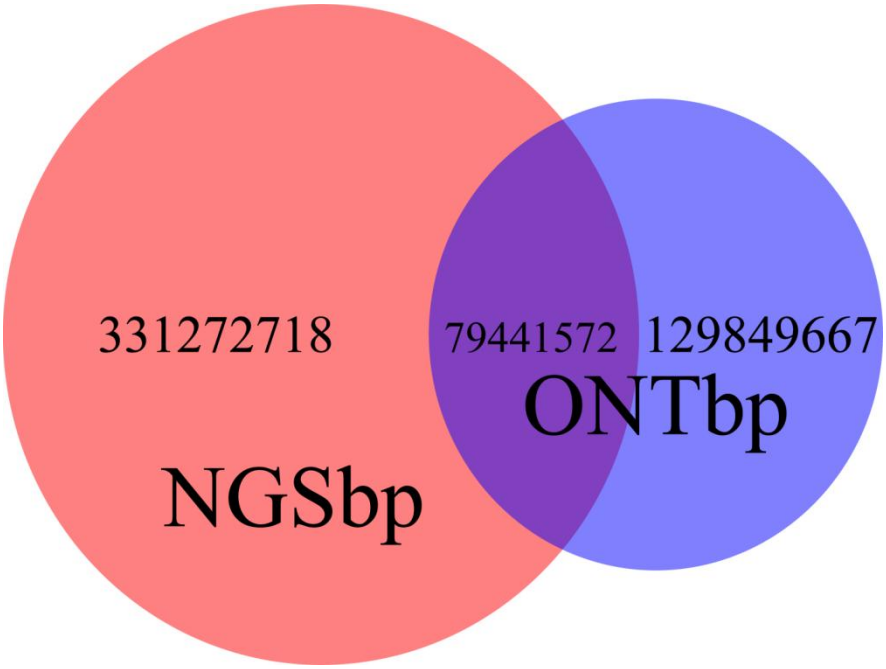

Figure S5. Base pair overlap of SVs between our study and the published datasets. Related to STAR Methods.

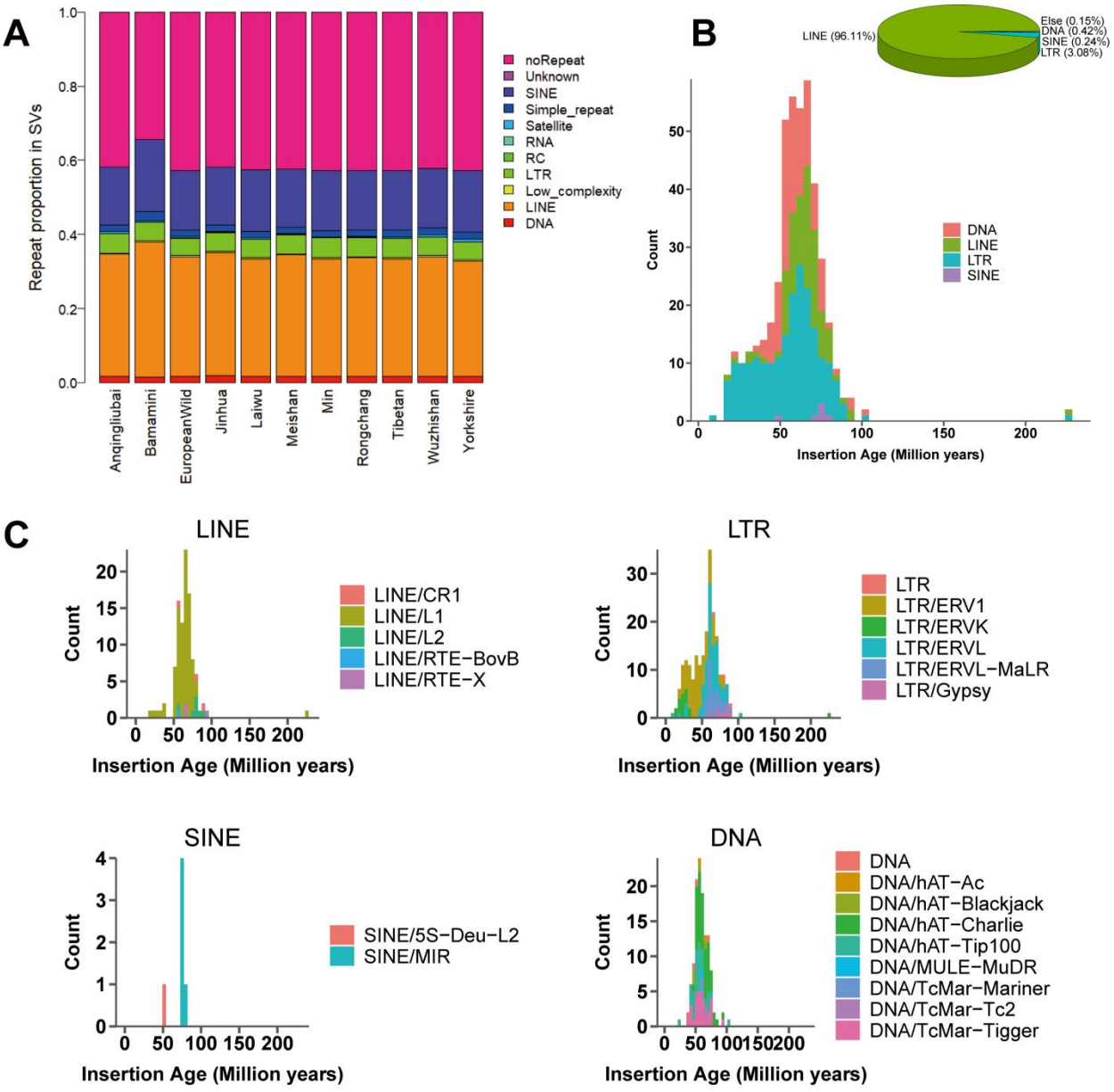

Figure S6. Repeat Annotation of SVs. Related to Figure 2 and STAR Methods. (A) Proportion of each repetitive sequence in the SV region per sample. SINE: short interspersed nuclear elements; LINE: long interspersed nuclear element; LTR: long tandem repeat; RC: rolling circles; noRepeat: Non-repetitive sequence. (B) The 3D pie chart shows the proportion of each class in the repeated sequences of the INS sequences. And the stacking plot shows the insertion age distribution of the four main superfamilies. (C) Insertion Age distribution of families within each superfamilies.

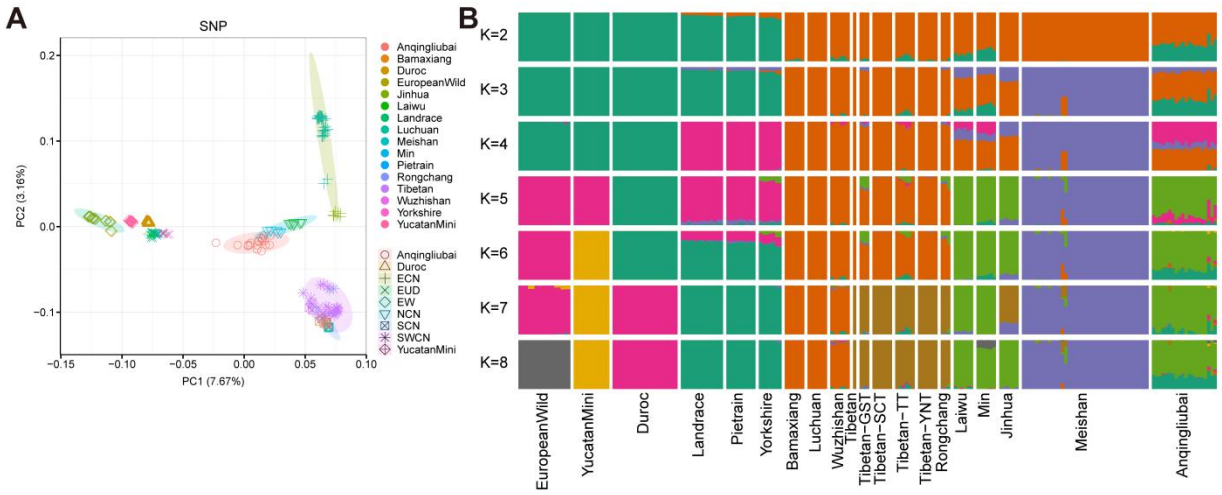

Figure S7. SNPs-based population structure. Related to Figure 3 and STAR Methods. (A) Principal component analysis. (B) Admixture analysis. See Table S2 for more details.

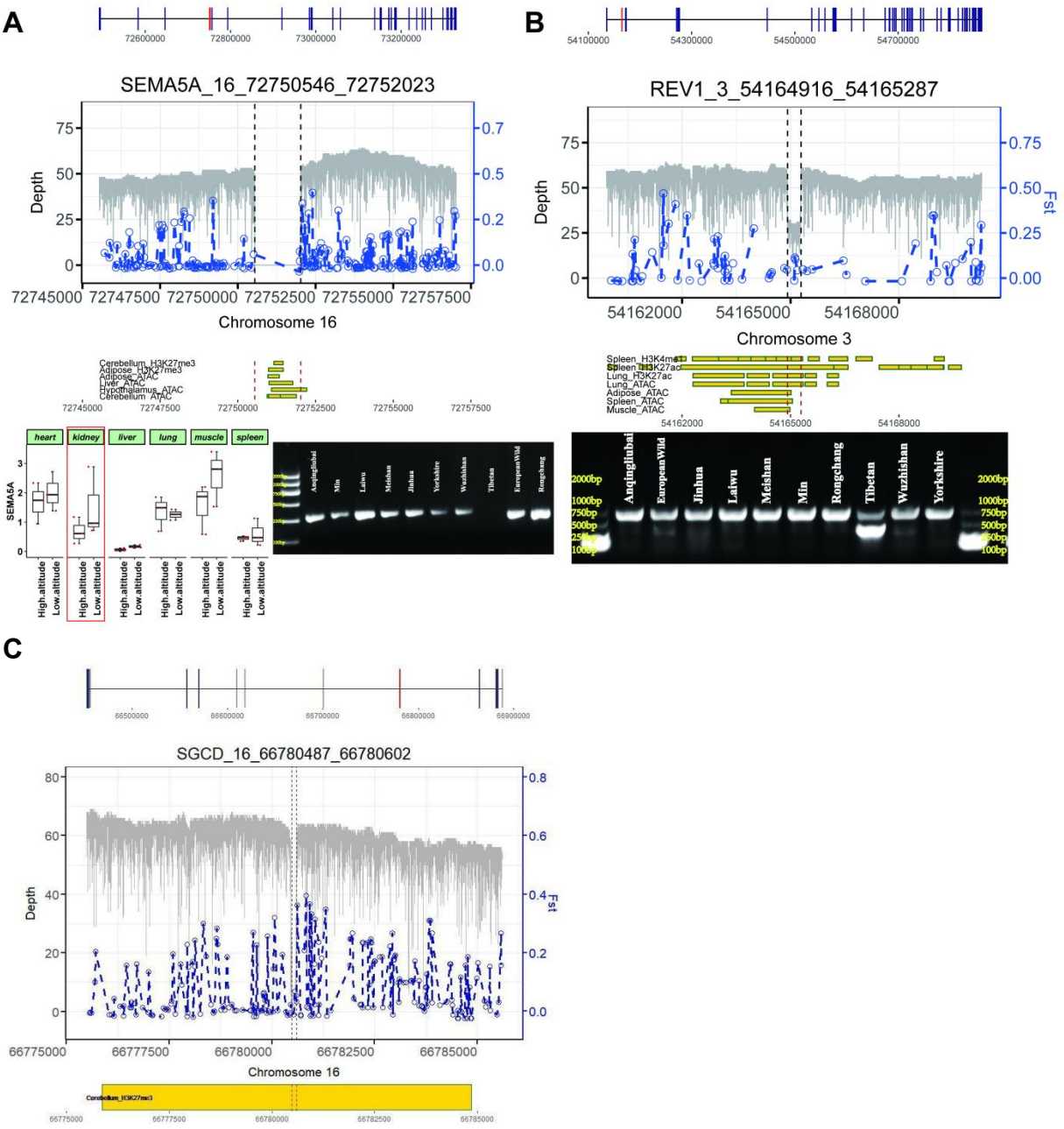

Figure S8. SVs of *SEMA5A* (A), *REV1* (B) and *SGCD* (C). Related to Figure 6. The upper panel shows a schematic map of the gene structure and SV location. The middle distribution showed the 5-kb regions upstream and downstream of this SV and the Fst statistic (Tibetan vs. low-altitude pigs) of SNPs in this region, as demonstrated in the left (gray) and right coordinate axes (blue), respectively. The corresponding epigenetic signals in this region are then shown. The left side of the bottom panel shows the gene expression value in corresponding tissues. The right side of the bottom panel shows the electromorphism of the SV.

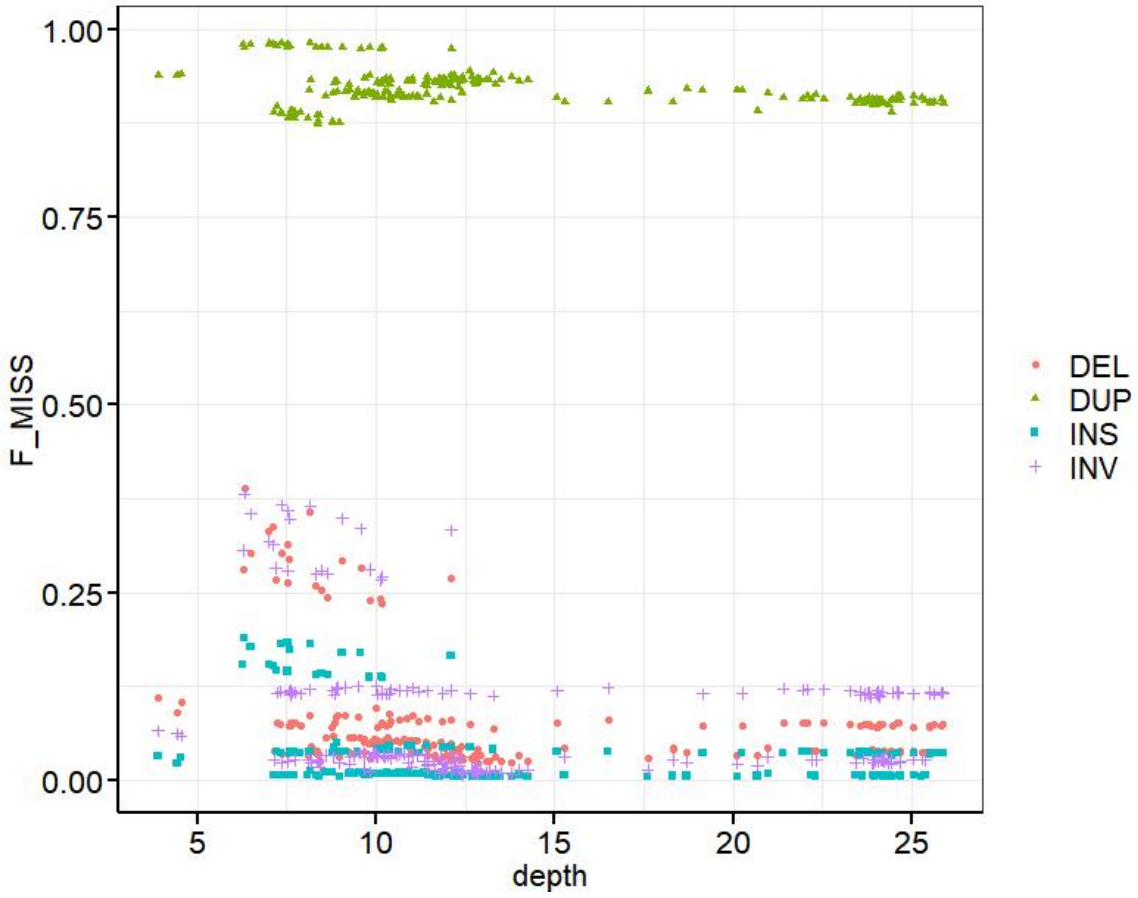

Figure S9. The genotyping missing rate varied with depth per individual. Related to STAR Methods.

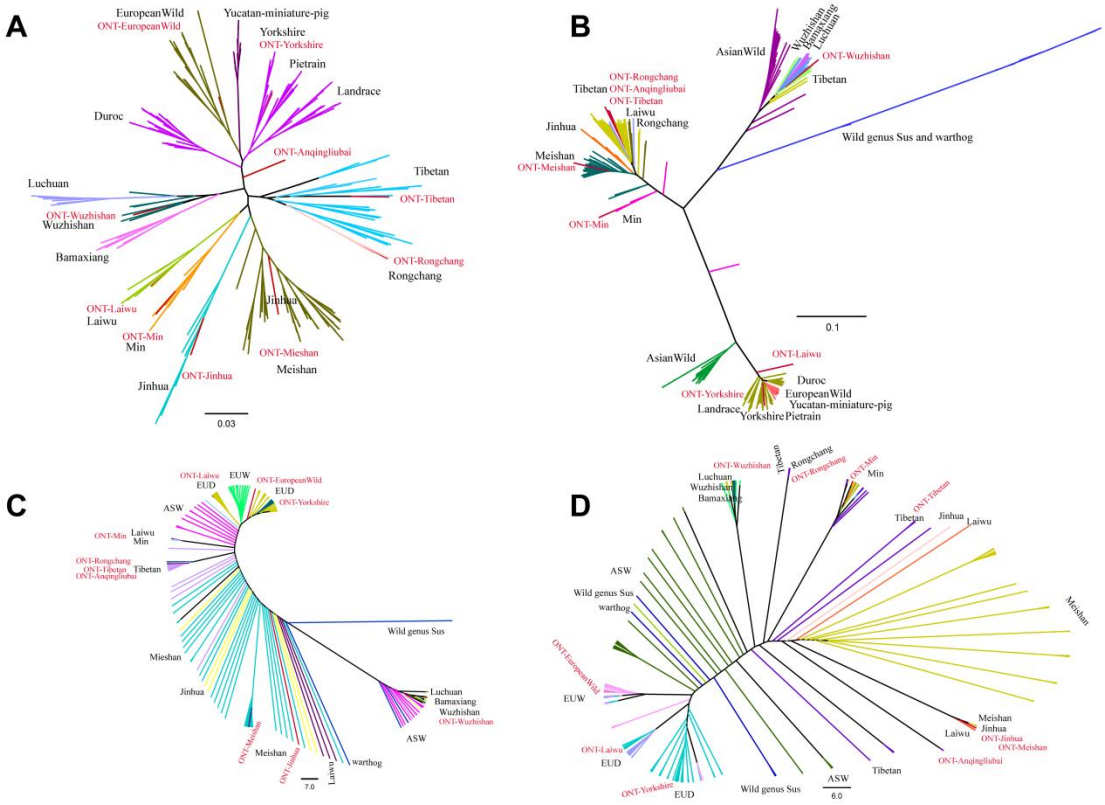

Figure S10. Phylogenetic tree. Related to Figure 4. The phylogenetic trees for different genomic regions, including autosome (A), chromosome X (B), 13-Mb (chrX:44-57Mb) region on chromosome X (C) and 30-Mb (chrX:57-87Mb) region on chromosome X (D).

Table S5: Annotation of repeat sequences in SVs. Related to Figure 2.

Class	SVs (bp)	Nonredundant SVs(bp)	Prop
DNA	5,859,781	299,563,963	1.96%
LINE	91,821,126		30.65%
Low_complexity	825,733		0.28%
LTR	15,089,179		5.04%
RC	27,172		0.01%
RNA	63,450		0.02%
Satellite	998,208		0.33%
Simple_repeat	4,682,036		1.56%
SINE	46,153,113		15.41%
Unknown	55,192		0.02%

SINE: short interspersed nuclear elements; LINE: long interspersed nuclear element; LTR: long tandem repeat; RC: rolling circles

Table S14. Primers used for PCR validation. Related to Figures 6 and S8.

gene	SVID	Types	Sequences (5'→3')
SOD1	13:195324807-195324931:DEL:ID58231	Forward	CCGCAGCGTCATGGATACTA
		Reverse	ACCCGAAGCCTCGTTAAACC
REV1	3:54164916-54165287:DEL:ID14139	Forward	AGCCCCAGTAGTGTCATCCT
		Reverse	TCAGTTCCACTGAGCGTCTTC
SEMA5A	16:72750546-72752023:DEL:ID70146	Forward	ATGTCCCATCGATCATTCCC
		Reverse	CATCCAAACACTCATATTTC