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.

Α	C. P. LEW. C. LL. C. M. M. M. M. M. M. C. L. C. LLEY. M. L. LWING, M.M. R. M. L. LEWING,	В		Contraction and the second
chr1		chr1	THE REAL PROPERTY AND	
chr2	A series of the ser	chr2	Hereith Reserves Anti- an anno an ann an an ann anns anns anns	
chr3	inden differences a la disentación estre en enteriori en estado de	chr3	al anter a children and a black a tribu with a child a transmitter of the mass of the	
chr4	the second s	chr4	fatie in an even administration contaction and the second of the	
chr5	tar faire a laine a Chaine an t-are sub article and an	chr5	tenerinden af dette an att Benerin and Andrea tabe beiden er tabe bei	
chr6	alena die denied in beland taal van die van die date, gaar is name taan die	chr6	person die andre en stande wellten in Willen von die die Lander versichen die Under Alexander versichen welchen	
chr7	verbentiganiller uns stadan errasiantilet anala bereiter afen.	chr7	relitant destrikes nor namber established with the said and in terrein series.	
chr8	The analysis was the first of the state of the	chr8	Anter and an experimental and a set of a set of the set	
chr9	Construct a manufacture and the advertision of the second and the advertise	chr9	Second & explore and an alteral science being work of an antiper-	
chr10	the of states in a finite business in a state state	chr10	Version and the Bass Branchistory and	INS
chr11	and the second se	chr11	and the second se	DEL
chr12	tering and the second	chr12	e served a de sites that the site of the test	INV
chr13	An and a second to make the make a second of the second of the second of the second as the second of the	chr13	Restant also and a sector of the sector of the sector of the sector of the sector and sector of the sector of the	
chr14	and some markets in a constant is according to the statement of the statem	chr14	alone with the Wildow State between all along and a state of the second state	
chr15	and a second statements in the statement was not an exact statement of	chr15	ning, Berning and States, States, States, and States, and States and States on States, and	
chr16	and an and a second	chr16	district former when a few the state of the	
chr17	and and a second se	chr17	interference of the second	
chr18	er freit zu mit eine auf eine	chr18	berden der sein dem soner einer	
chrX	an a	chrX	area a da sincia a contrata a contrata da contrata da contrata da contrata da contrata da contrata da contrata	

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.





Proportion of each repetitive sequence in the SV region per sample. SINE: short interspersed nuclear element; 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 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).

Class	SVs (bp)	Nonredundant SVs(bp)	Prop
DNA	5,859,781		1.96%
LINE	91,821,126 omplexity 825,733		30.65%
Low_complexity		0.28%	
LTR	15,089,179	200 562 062	5.04%
RC	27,172		0.01%
RNA	63,450	299,303,903	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%

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

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

gene	SVID	Types	Sequences (5'→3')
SOD1	13:195324807-	Forward	CCGCAGCGTCATGGATACTA
SODI	195324931:DEL:ID58231	Reverse	ACCCGAAGCCTCGTTAAACC
DEV1	3:54164916-	Forward	AGCCCCAGTAGTGTCATCCT
KE V I	54165287:DEL:ID14139	Reverse	TCAGTTCCACTGAGCGTCTTC
SEMA5A	16:72750546-	Forward	ATGTCCCATCGATCATTCCC
SEMAJA	72752023:DEL:ID70146	Reverse	CATCCAAACACTCATATTTC

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