

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/189890/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Chen, Jian-Hai, Du, Xiaoyong, Zheng, Zhuqing, Gorkhali, Neena Amatya, Xiao, Changyi and Stanton, David W. G. 2026. Ancient introgression drives wild boar expansion and phenotypic diversification of domestic pigs. *Science* 393 (6818) , pp. 1335-1341. 10.1126/science.adq7553

Publishers page: <https://doi.org/10.1126/science.adq7553>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Title: Ancient introgression drives wild boar expansion and phenotypic diversification of domestic pigs

Authors: Jian-Hai Chen^{1,2,4,†,‡}, Xiaoyong Du^{1,3,44,†}, Zhuqing Zheng^{1,2,†}, Neena Amatyia Gorkhali^{5,†}, Changyi Xiao³, Shilin Zhu¹, Jie Zhong⁶, Liangliang Fu^{1,2}, Jingjin Li^{1,2}, Zhuofan Yan³, Edward S. Ricemeyer^{7,8}, Fan Jiang⁹, Thuy Nhien Tran Thi^{1,2}, Pan Ni¹, Nicholas VanKuren⁴, Guanglin He⁶, Yalan Yang¹⁰, Supamit Mekchay¹¹, San San Hmwe¹², Mikhail Sablin¹³, Nikica Šprem¹⁴, Le Thi Thuy¹⁵, Abdullah Omar¹⁶, M.O. Faruque¹⁷, Ali Akbar Bhuiyan¹⁸, G. L. L. Pradeepa Silva¹⁹, Amali Malshani Samaraweera²⁰, Debabrata Phukon²¹, Mukesh Thakur²², Saroj Sapkota⁵, S. M. Farhad Vahidi^{23,24}, Faiz-ul Hussan²⁵, Hidayat Ashari^{26,27}, Bayu Dewantoro Putro Soewandi^{1,27}, Ivan Jakovlić²⁸, Kevin G. Daly^{29,30}, Ming-Shan Wang³¹, Xiaolei Liu^{1,2}, Malavanh Chittavong³², Daniel Falush³³, Jeffrey Spence^{34,35}, Richard Bishop⁴², David W. G. Stanton^{36,37}, Xuewen Xu^{1,2}, Xinyun Li^{1,2}, Matthew W. Hahn^{38,39}, Lingzhao Fang⁴⁰, Laurent A. F. Frantz^{7,36*}, Greger Larson^{41*}, Jianlin Han^{42,43,44*}, Shuhong Zhao^{1,2,44*}

¹Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University; Wuhan, PR China.

²Frontiers Science Center for Animal Breeding and Sustainable Production; Wuhan, PR China.

³Key Laboratory of Intelligent Technology in Animal Husbandry, Ministry of Agriculture and Rural Affairs, Engineering Research Center of Smart Agricultural Technology, Ministry of Education, Huazhong Agricultural University; Wuhan, PR China.

⁴Department of Ecology and Evolution, The University of Chicago; Chicago, IL, United States.

⁵National Animal Breeding and Genetics Research Centre, Nepal Agriculture Research Council; Khumaltar, Nepal.

⁶Institute of Rare Diseases, West China Hospital of Sichuan University, Sichuan University; Chengdu, PR China.

⁷Chair of Animal Systems Genomics, Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany.

⁸Bond Life Sciences Center, University of Missouri; Columbia, MO, USA.

⁹Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences, Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province; Wuhan, PR China.

¹⁰Department of Human Genetics, The University of Chicago; Chicago, IL, USA.

¹¹Department of Animal and Aquatic Sciences, Faculty of Agriculture, Chiang Mai University; Chiang Mai, Thailand.

¹²Dagon University; Yangon, Myanmar.

¹³Zoological Institute of the Russian Academy of Sciences; St. Petersburg, Russia.

¹⁴University of Zagreb Faculty of Agriculture, Department of Fisheries, Apiculture, Wildlife Management and Special Zoology; Zagreb, Croatia.

¹⁵National Institute of Animal Sciences; Hanoi, Vietnam.

¹⁶National Engineering Laboratory for Animal Breeding, Key Lab of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, China Agricultural University; Beijing, China.

¹⁷Department of Animal Breeding and Genetics, Bangladesh Agricultural University; Mymensingh, Bangladesh.

¹⁸Livestock Division, Bangladesh Agricultural Research Council (BARC); Dhaka, Bangladesh.

¹⁹Department of Animal Science, Faculty of Agriculture, University of Peradeniya; Peradeniya, Sri Lanka.

²⁰Department of Animal Science, Faculty of Animal Science and Export Agriculture, Uva Wellasa University; Badulla, Sri Lanka.

²¹Animal Husbandry and Veterinary Department, Government of Assam; Guwahati, India.

²²Mammal and Osteology Section, Zoological Survey of India; Kolkata, India.

²³Animal Science Research Department, Qom Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO); Qom, Iran.

²⁴Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO); Karaj, Iran.

²⁵Department of Breeding and Genetics, Cholistan University of Veterinary and Animal Sciences; Bahawalpur, Punjab, Pakistan.

²⁶National Research and Innovation Agency (BRIN); Jl. Raya Jakarta Bogor Km 46, Cibinong, West Jawa, Indonesia.

²⁷Research Organization for Agriculture and Food, National Research and Innovation Agency (BRIN); Jl. Raya Jakarta-Bogor KM. 46, Cibinong 16911, Jawa Barat, Indonesia.

²⁸State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, and College of Ecology, Lanzhou University; Lanzhou, PR China.

²⁹Smurfit Institute of Genetics, Trinity College Dublin; Dublin, Ireland.

³⁰School of Agriculture and Food Science, University College Dublin; Dublin, Ireland.

³¹Kunming Institute of Zoology, Chinese Academy of Sciences; Kunming, PR China.

³²Department of Livestock and Fisheries, Faculty of Agriculture, National University of Laos; Vientiane, Lao People's Democratic Republic.

³³Institute of Materia Medica, Chinese Academy of Sciences; Shanghai, PR China.

³⁴Institute for Human Genetics, University of California, San Francisco; San Francisco, CA, USA.

³⁵Department of Epidemiology and Biostatistics, University of California, San Francisco; San Francisco, CA, USA.

³⁶School of Biological and Behavioural Sciences, Queen Mary University of London; London, UK.

³⁷Cardiff School of Biosciences, Cardiff University; Cardiff, UK.

³⁸Department of Biology, Indiana University; Bloomington, IN, USA.

³⁹Department of Computer Science, Indiana University; Bloomington, IN, USA.

⁴⁰Center for Quantitative Genetics and Genomics (QGG), Aarhus University; Aarhus, Denmark.

⁴¹PalaeoBARN, School of Archaeology, University of Oxford; Oxford, UK.

⁴²International Livestock Research Institute (ILRI); Nairobi, Kenya.

⁴³CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS); Beijing, PR China.

⁴⁴Yazhouwan National Laboratory; Sanya, Hainan, PR China.

‡Present address: University of California, Irvine; Irvine, CA 92697, USA.

†These authors contributed equally: Jian-Hai Chen, Xiaoyong Du, Zhuqing Zheng, and Neena Amatya Gorkhali

*Corresponding author: Laurent.Frantz@palaeo.vetmed.uni-muenchen.de; greger.larson@arch.ox.ac.uk; hanjianlin@yzwlab.cn; zhaoshuhong@yzwlab.cn.

Abstract: Despite the ubiquity of wild boar populations across Eurasia, their origins, routes of dispersal, and phenotypic evolution remain uncertain. Here, using 745 genomes representing populations across Eurasia, we demonstrate that wild boar expanded from South Asia during the Middle Pleistocene. This expansion was preceded by a Pliocene introgression (~3–5 Mya) in South Asia from an archaic *Suinae* lineage that introduced a ~43 Mb X-chromosomal haplotype into the ancestral wild boar population. This haplotype includes loci associated with increased body size and adaptation to cold climates. During domestication, this region was selected at least three times independently to develop modern large-bodied breeds. Our findings reveal a long-lasting evolutionary legacy of this ancient introgression, which provided genetic material later shaped by natural and artificial selection.

Main Text: Wild boar (*Sus scrofa*) is among the most widely distributed terrestrial mammals. Their populations occupy diverse habitats across Eurasia and North Africa including tropical forests and the subarctic (1-3). Although fossil and genetic evidence suggests a Pleistocene expansion from tropical Asia (~1–0.5 Mya) (4-6), sparse genomic sampling from South Asia has left unresolved whether South Asia, mainland Southeast Asia (MSEA), or island Southeast Asia (ISEA) harbors the deepest-diverging wild boar lineage.

The adaptation of wild boar across diverse ecosystems during their expansion was likely facilitated by interspecific introgression (5, 7). Subsequent independent pig domestication processes in Western and Eastern Eurasia (1), together with gene flow between wild boar and domestic pigs (8), possibly favored introgressed variants to promote phenotypic diversity in pigs. One highly variable phenotype important to domestication in pigs is body size (9). Tropical Asian miniature breeds, widely used as biomedical models (10), weigh only ~20 kg, whereas some European large-bodied breeds exceed 400 kg (11, 12). In wild boar, variation in body size often follows Bergmann's rule, where larger individuals occupy colder climates (13). An archaic X-chromosomal introgression (~40 Mb) has been implicated in cold adaptation of Western and Northern Eurasian wild boar (7), yet its evolutionary origin and relevance to body-size diversification remain unclear.

Here, we analyzed 745 genomes from 58 populations worldwide (table S1; see materials and methods), including three de novo assemblies and 256 newly sequenced samples from South Asia and MSEA (fig. S1; tables S2 to S4). Using population genomic approaches, we assessed the phylogeographic origin of wild boar, mapped the evolutionary trajectory of the ancient introgression, tested its contributions to body-size variation and climate adaptation, and evaluated its role in selection during pig domestication.

South Asian wild boar as a sister lineage to other wild boar and domestic pigs

Phylogenetic trees of mitogenomes, Y chromosomal sequences, and genome assembly k-mers consistently indicated that South Asian wild boar represent a distinct lineage sister to all other wild boar and domestic pig lineages (100% bootstrap support; figs. S2 to S3; tables S5 to S7). This model of South-Asian wild boar as sister group to all other groups (hereafter the South-Asia-sister topology) was also favored by admixture graphs, i.e., qpGraph (14) ($|Z| = 1.75$, fig. S4) and AdmixtureBayes (15) (posterior probability = 0.896; **Fig. 1A**), whereas alternative models involving interspecific introgression into South Asian wild boar received negligible support (posterior probability = 0.0001; fig. S5). PCA further revealed a clear genetic differentiation of South Asian wild boar and domestic pigs from other groups (**Fig. 1B** and fig. S6; table S1).

The South Asian group exhibited higher historical effective population size (N_e), greater genetic diversity, faster linkage disequilibrium (LD) decay, and a lower inbreeding coefficient (**Fig. 1C** and figs. S7 to S9; table S8). Of the ~104.78 million SNPs genotyped in our dataset, 45.47% (47.64 million) were novel and unreported in the dbSNP database (16). Of these new SNPs, 70.29% were common (MAF > 0.05), and 14.19% were private to South Asia, 3.59-fold more than all other groups combined (table S9), arguing against recent admixture as the main cause of high South Asian diversity.

To assess the frequency of the South-Asia-sister topology, we applied topology weighting with Twisst (17), which showed that this topology was supported more often than the ISEA- or MSEA-sister alternatives across sliding windows (**Fig. 1D**; table S10). This inference was robust to local non-monophyly: even among non-monophyletic windows, the majority placed South

Asian samples sister to the remaining sampled *S. scrofa* populations (63.91% of autosomal windows and 73.49% of X-linked windows; table S11). The ISEA-sister signal likely reflects interspecific introgression from *S. verrucosus* rather than incomplete lineage sorting (ILS), as supported by concordant evidence from D-statistics (18), TreeMix (19), AdmixtureBayes, QuIBL (20), qpGraph (14), and f4-ratio analyses (21) (**Fig. 1A** and figs. S10 to S13), as well as prior studies (22). Using demographic modeling with Momi2 (23), we estimated introgression from *S. verrucosus* into ISEA wild boar at ~325 Kya (303–341 Kya) (fig. S14).

Because high recombination can facilitate the persistence of introgressed ancestry (24, 25), we explored its association with the ISEA-sister topology. Fine-scale recombination rate was positively correlated with both the Twisst weight for this topology and *S. verrucosus*–ISEA allele sharing ($A_{Sver-ISEA}$) (figs. S15 to S17; tables S12 and S13). After controlling for GC content, CpG density, local mutation-rate proxy, and mappability, the correlation between recombination rate and $A_{Sver-ISEA}$ remained moderate to strong on autosomes (ρ : 0.637–0.983) and the X chromosome (ρ : 0.562–0.616) (fig. S17).

In contrast, regions supporting the South-Asia-sister topology exhibited significantly lower recombination rates than those supporting the MSEA- or ISEA-sister alternatives (Wilcoxon rank-sum test, $P < 0.0001$; **Fig. 1E**), a pattern consistent with the expectation that low-recombination regions may be less affected by complex reticulate histories (26). To test whether this signal reflected recombination-rate bias, we repeated AdmixtureBayes after excluding the lowest or highest 20% recombination-rate bins. In both analyses, the South-Asia-sister topology remained the top-ranked minimal topology (PP = 0.780 and 0.707; fig. S18), indicating that this topology was not driven by recombination-rate extremes. These findings support the South-Asia-sister topology as the predominant evolutionary history of wild boar, whereas the ISEA-sister signal appears partly shaped by *S. verrucosus* introgression and its preferential retention in high-recombination regions.

Wild boar divergence times and domestic pig ancestry

To estimate divergence times among wild boar populations, we used autosomal segments supporting the South-Asia-sister topology to minimize the impact of interspecific introgression. MSMC2 analysis (27, 28) estimated the South Asian wild boar separation from other lineages at 658–695 Kya, substantially predating the divergence among all pairs of non-South Asian populations (113–360 Kya; **Fig. 1F** and fig. S19).

Pigs were independently domesticated from wild boar in Asia and Europe (1). Crossbreeding between the two domestic populations since the 18th century has created mixed ancestry in some breeds (8, 29, 30). To quantify European and Asian ancestry across South and Southeast Asian populations, we performed haplotype-based admixture analyses (31, 32). We detected European admixture in some local populations, such as Nepalese Hurrah (31.44%), Thai (28.78%), and Myanmar pigs (75.17%) (fig. S20; tables S14 and S15), with most admixture events (92.3%, $n = 13$) occurring within the past ~73 generations. In contrast, several local breeds, including Laotian Small-Eared, Eastern Nepalese Bampudke, and plainland Bangladeshi pigs, showed no detectable European admixture.

Mitogenomic and Y-chromosomal analyses revealed predominantly Asian maternal (69% South, 31% East/Southeast Asian) and mixed paternal origins (75% Asian, 25% European) (fig. S2; tables S5 and S6). Given the larger body size of European commercial pigs than Asian

indigenous pigs, their genetic contribution is consistent with recent crossbreeding aimed at increasing body size (9).

South Asian origin of the large X-chromosomal introgressed haplotype

The deep divergence of the South Asian lineage suggests that it may be the source of the highly divergent X-chromosomal haplotype previously identified in European and northern Chinese wild boar (7, 33). We tested this hypothesis using *Saguaro* (34) and *Twisst* (17) on the X chromosome. A large continuous region (~44–87 Mb) distinguished South Asian large-bodied populations (e.g., Pakistani wild boar and Bangladeshi pigs) from miniature ones (e.g., Nepalese wild boar and domestic pigs) through a distinct phylogenetic topology (**Fig. 2A** and fig. S21). This pattern was consistently observed across global populations: large-bodied *S. scrofa* populations from South Asia, Europe, and northern China formed a sister lineage to a clade comprising pygmy hog, other *Sus* species, and miniature *S. scrofa* from tropical Asian regions, with a divergence time of ~8.70–7.01 Mya (**Fig. 2B** and fig. S22).

We further tested whether the large-bodied associated haplotype was introduced into the population through introgression with an extinct Suinae lineage (7, 33). Archaic segment detection using *Sprime* (35) and *HMMix* (36) provided support for ancient introgression, with enrichment concentrated in the 44–87 Mb region (*Sprime* scores > 95th percentile; *HMMix* posterior probability > 80%; fig. S23). A permutation test showed that this enrichment was robust to different outgroup choices and masking strategies (combined $Z = 4.16$, empirical $P = 0.030$; fig. S24). Coalescent analysis based on ancestral recombination graphs using *SINGER* (37) yielded a deeper time to the most recent common ancestor (TMRCA) for the ~44–87 Mb region than for autosomal background (fig. S25), consistent with acquisition of the large-bodied associated haplotype from a divergent lineage. In addition, coalescent simulations with *msprime* (38) indicated that the observed TMRCA deviated markedly from ILS-only expectations ($P < 0.001$; fig. S26), suggesting that the depth of coalescence cannot be explained by ILS alone.

Within this introgressed region, two subregions (~44–57 Mb and ~57–87 Mb) exhibited distinct evolutionary histories in northern Chinese domestic pigs (**Fig. 2B** and fig. S21). However, both subregions positioned South Asian wild boar sister to the European and northern Asian wild boar lineages, with divergence times of ~5.45–3.05 Mya. Combined with the broad geographic distribution of this South Asian lineage across the Ganges–Brahmaputra Delta, the Indus River basin, and Sri Lanka, these results support a South Asian origin of the introgressed haplotype within *S. scrofa* (**Fig. 2, C and D**).

Association between the X-chromosomal introgression and body size

Marked divergence in the introgressed region (44–87 Mb) between miniature and large-bodied pigs suggests an association with body size. To test this hypothesis, we performed a genome-wide association study (GWAS) for adult weight/breed-level body size using a linear mixed model with an autosomal relatedness matrix, sex, and PCs to account for population structure ($\lambda_{\text{autosome}} = 1.010$; **Fig. 3A** and fig. S27; table S16). Nearly all significant SNPs mapped to the 44–87 Mb region (16,365/16,383; 99.9%), with the strongest association signal concentrated in the first subregion (~44–57 Mb; Bonferroni $P < 2.04 \times 10^{-9}$). LD analysis revealed uniformly high D' (mean 0.963) but heterogeneous R^2 among variants (mean 0.595; fig. S28), suggesting limited recombination across this region alongside multiple associated variants with variable allele frequencies. Coalescent simulations with *msprime* (38) showed that this LD pattern can be recapitulated under strong positive selection ($s = 0.02$) and reduced recombination (~10² fold) (fig. S29).

Archaeological observations in wild boar link larger body size to colder environments, consistent with Bergmann's rule (13). We therefore tested whether body-size-associated genetic variation was also associated with climatic variables, using 72 bioclimatic variables from the Last Glacial Maximum (~21 Kya) and recent decades (1979–2013) ($\lambda = 0.992$ – 1.036 ; table S17). Sixteen climate variables representing winter-to-spring minimum temperature and precipitation showed strong X-chromosomal enrichment, with >85% of the top 100 association hits within the 44–57 Mb subregion (**Fig. 3B** and figs. S27 and S30; $P < 1.97 \times 10^{-9}$). Bidirectional Mendelian randomization (MR) (39) provided evidence consistent with a causal contribution of larger body size to adaptation to colder environments (IVW method β : -0.21 to -0.11, FDR < 0.05), whereas reverse environmental effects were not significant (**Fig. 3C**). Estimates were consistent across three additional MR algorithms (e.g., minimum February temperature IVW $\beta = -0.21$, MR-PRESSO $\beta = -0.19$; fig. S31), with no detectable directional pleiotropy by MR-Egger regression (table S18). Collectively, this ancient introgression likely contributed to body-size variation that facilitated wild boar adaptation to diverse Pleistocene environments.

Regulatory associations of the X-chromosomal introgressed haplotype

Despite strong DNA-level LD across the large introgressed region (44–87 Mb), allele-specific expression (ASE) can provide gene-level regulatory readouts less directly confounded by long-range LD (40). To test this, we used an F₁ resource population generated from large-bodied Duroc sires $\times$ miniature Luchuan dams ($n = 189$, 77 females; **Fig. 3D**; table S19) and analyzed the association between ASE of muscle-expressed genes and carcass straight length, a representative adult body-size trait. Genome-wide comparison showed that this correlation peaked in the 44–57 Mb subregion, with opposite directions for the two haplotypes (median $r = 0.20$ and -0.20 for the large-bodied and miniature haplotypes, respectively; both $P < 10^{-5}$ versus autosomal background, Mann–Whitney U test; figs. S32 and S33), suggesting a muscle-specific regulatory association with adult body-size-related traits.

To test whether this regulatory signal extended beyond carcass length, we performed a phenome-wide association study (PheWAS) across 11 growth- and carcass-related traits in the F₁ population, including weights at three stages, body length, eye muscle area, backfat thickness, and vertebral and rib counts. Multiple body-size traits were significantly associated with ASE of 35 genes in this region (Bonferroni $P < 0.004$, **Fig. 3E**; table S20). Postnatal growth traits, such as carcass length and slaughter weight, were mainly associated with genes preferentially expressed from the large-bodied haplotype, whereas prenatal and early developmental traits, such as vertebral and rib counts, birth weight, and weaning weight, were linked to genes preferentially expressed from the miniature haplotype. Among these associations, 86% (12/14) involving large-bodied haplotype-biased expression correlated positively with body-size traits (e.g., *ZC4H2*, *EFNB1*, *MAGED1*), whereas 68% (23/34) involving miniature haplotype-biased expression showed negative correlations (e.g., *AMER1*, *SMC1A*, *GLA*). Several of these genes are annotated for development and growth functions (41) (table S21), supporting context-specific regulatory associations between this region and body-size-related traits.

To characterize the cis-regulatory architecture underlying these ASE patterns, we analyzed enhancer enrichment in the 44–87 Mb region across 14 major pig tissues using integrative chromatin-profiling datasets, including ChIP-seq for histone modifications and ATAC-seq (42, 43). We identified six enhancer-enriched candidate regulatory subregions (0.5–1.0 Mb in width; 44.0–44.5, 56.0–56.5, 57.5–58.0, 63.0–63.5, 83.0–83.5, and 84.5–85.0 Mb) with moderate to strong enhancer enrichment (**Fig. 3F** and fig. S34; tables S22 and S23). These subregions showed distinct tissue enrichment profiles, including the central nervous system and muscle

(~44.0 Mb), gastrointestinal tract (~63.0 Mb), neuroendocrine system (~84.5 Mb), and multiple metabolic and digestive organs (~57.5–58.0 Mb). These chromatin annotations provide supportive evidence that enhancer-enriched subregions may contribute to ASE-associated regulatory architecture, including potential contributions from transposable-element-derived enhancers (44).

Historical recombination suppression between haplotypes and structural variant accumulation

In the ~44–57 Mb subregion, South Asian, northern Chinese, and European pigs (medium- to large-bodied) carry the introgressed haplotype, whereas miniature breeds carry the ancestral-like haplotype, which is also represented in other extant *Sus* species and pygmy hog (**Figs. 4A and 2B**). The deep genetic divergence between these two haplotypes in this subregion (**Figs. 4B, 2B**, and fig. S35) suggests reduced recombination between them. Consistent with this, historical recombination between the two divergent haplotypes was suppressed in populations carrying both haplotypes (fig. S36).

We then explored whether structural variants (SVs) had accumulated in this region. SVs were identified using SyRI (45) and validated using a pangenome graph-based approach (46). On the X chromosome, we detected 134 medium-to-large SVs (100 bp–77.5 kb), of which 97.01% (130/134) were associated with the large-bodied versus miniature haplotype groups (Benjamini–Hochberg FDR < 0.05; figs. S37 and S38; tables S24 and S25). Most of these SVs (126/134, 94.0%) were concentrated within the 44–57 Mb subregion, including most insertions (95.65%) and deletions (90.00%), as well as all inversions and highly divergent regions. Together, these SVs may have contributed to or reinforced recombination suppression between the introgressed and ancestral haplotypes (47).

Selection on variation within the X-chromosomal introgressed haplotype in domestic populations

Although recombination was suppressed between the ancestral and introgressed haplotypes, linkage maps from populations fixed for either haplotype showed recombination rates comparable to the rest of the X chromosome (fig. S36), suggesting that recombination can proceed normally between homologous copies of the same haplotype. Nevertheless, domestic populations carrying the introgressed haplotype showed longer-range LD and lower nucleotide diversity than wild boar (**Fig. 4C** and figs. S39 and S40). Together, these patterns are consistent with positive selection in domestic populations, likely favoring increased body size (9), acting on the limited standing genetic variation within the introgressed haplotype.

To further test this hypothesis, we computed several summary statistics, including F_{ST} , Tajima's D, allele frequency change (ΔAF), and extended haplotype homozygosity (EHH) decay. Within the 44–57 Mb subregion, F_{ST} was 3.0–3.9-fold higher than the genome-wide background across northern Chinese, European, and South Asian pigs (top 0.55%–1.38% genome-wide; **Fig. 4D** and fig. S41). An F_{ST} enrichment test showed that body-size-associated SNPs had significantly elevated population differentiation compared with MAF-matched control SNPs ($P < 0.0001$; fig. S42), consistent with genetic differentiation shaped by selection (48). Domestic pigs also exhibited more negative Tajima's D than wild boar populations (**Fig. 4E** and fig. S40).

Furthermore, consistent with polygenic selection (49), frequencies of body-size-increasing alleles (ΔAF for GWAS alleles with $\beta > 0$; $P < 2.04 \times 10^{-9}$) were higher in larger-bodied domestic pigs than in wild boar (Wilcoxon signed-rank test, $P < 5.92 \times 10^{-11}$; **Fig. 4F** and fig.

S40). Finally, consistent with the nucleotide diversity estimates, EHH profiles further confirmed reduced haplotypic diversity in domestic pigs across this subregion (fig. S43). Given the long-range haplotype structure and heterogeneous LD across this region, we interpreted these selection statistics as complementary haplotype-level evidence rather than independent SNP-level signals. Together with independent domestication histories in northern Chinese, European, and South Asian domestic pigs (*1*), these results suggest that selection on this introgressed region may have occurred at least three times independently.

Our results suggest that Eurasian wild boar derived from a South Asian lineage that expanded during the Middle Pleistocene. Prior to this dispersal, admixture (~3–5 Mya) with a distantly related Suinae lineage likely introduced a divergent ~43-Mb haplotype into ancestral wild boar in South Asia, which later spread to North Asia and Europe during the Pleistocene expansion. This haplotype is associated with increased body size in northern-latitude populations, consistent with a selective advantage in colder climates under Bergmann's rule (*13*). Over millions of years, additional mutations and structural variants accumulated within this extended X-chromosome region, establishing a reservoir of standing variation that was later reshaped by selection during domestication to generate modern large-bodied pig breeds. Collectively, these findings show how ancient introgression can provide a long-lasting reservoir of genetic variation for both natural and human-mediated selection.

Materials and Methods are available in the Supplementary Materials.

References and Notes

1. G. Larson *et al.*, Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618-1621 (2005).
2. J. Chen *et al.*, Selective constraints in cold-region wild boars may defuse the effects of small effective population size on molecular evolution of mitogenomes. *Ecology and Evolution* **8**, 8102-8114 (2018).
3. M. F. Rothschild, A. Ruvinsky, *The genetics of the pig*. (CABI, 2011).
4. M. A. M. Groenen *et al.*, Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398 (2012).
5. L. A. F. Frantz *et al.*, The evolution of Suidae. *Annual Review of Animal Biosciences* **4**, 61-85 (2016).
6. L. A. F. Frantz *et al.*, Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol.* **14**, R107 (2013).
7. H. Ai *et al.*, Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nature Genetics* **47**, 217-225 (2015).
8. L. A. F. Frantz *et al.*, Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature Genetics* **47**, 1141-1148 (2015).
9. C. Lega, P. Raia, L. Rook, D. Fulgione, Size matters: A comparative analysis of pig domestication. *The Holocene* **26**, 327-332 (2016).
10. J. K. Lunney *et al.*, Importance of the pig as a human biomedical model. *Science Translational Medicine* **13**, eabd5758 (2021).
11. K. Nidup, D. D. Joshi, J. Gongora, C. Moran, Farming and biodiversity of indigenous pigs in Nepal. *Biodiversity* **11**, 26-33 (2010).
12. H. M. Briggs, *Modern breeds of livestock*. (Macmillan, 1969).
13. U. Albarella, K. Dobney, P. Rowley-Conwy, Size and shape of the Eurasian wild boar (*Sus scrofa*), with a view to the reconstruction of its Holocene history. *Environmental Archaeology* **14**, 103-136 (2009).
14. R. Maier *et al.*, On the limits of fitting complex models of population history to f-statistics. *eLife* **12**, e85492 (2023).
15. S. V. Nielsen *et al.*, Bayesian inference of admixture graphs on Native American and Arctic populations. *PLoS genetics* **19**, e1010410 (2023).
16. S. T. Sherry *et al.*, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308-311 (2001).

17. S. H. Martin, S. M. Van Belleghem, Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics* **206**, 429-438 (2017).
18. N. Patterson *et al.*, Ancient Admixture in Human History. *Genetics* **192**, 1065-1093 (2012).
19. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**, e1002967 (2012).
20. N. B. Edelman *et al.*, Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594-599 (2019).
21. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources* **21**, 584-595 (2021).
22. L. A. Frantz, O. Madsen, H. J. Megens, M. A. Groenen, K. Lohse, Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian Sus species during the Plio-Pleistocene climatic fluctuations. *Molecular Ecology* **23**, 5566-5574 (2014).
23. J. Kamm, J. Terhorst, R. Durbin, Y. S. Song, Efficiently inferring the demographic history of many populations with allele count data. *Journal of the American Statistical Association* **115**, 1472-1487 (2020).
24. M. Schumer *et al.*, Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656-660 (2018).
25. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol* **17**, e2006288 (2019).
26. J. B. Pease, M. W. Hahn, More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* **67**, 2376-2384 (2013).
27. M. Zhang, Q. Yang, H. Ai, L. Huang, Revisiting the evolutionary history of pigs via De Novo mutation rate estimation in a three-generation pedigree. *Genomics, Proteomics & Bioinformatics*, 13 (2022).
28. S. Schiffels, R. Durbin, Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46**, 919-925 (2014).
29. M. Bosse *et al.*, Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nature Communications* **5**, 4392 (2014).
30. B. Yang *et al.*, Genome-wide SNP data unveils the globalization of domesticated pigs. *Genetics Selection Evolution* **49**, 71 (2017).
31. G. Hellenthal *et al.*, A genetic atlas of human admixture history. *Science* **343**, 747-751 (2014).
32. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, Inference of population structure using dense haplotype data. *PLoS Genetics* **8**, e1002453 (2012).
33. L. Liu *et al.*, Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nature Communications* **10**, 1992 (2019).
34. N. Zamani *et al.*, Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics* **14**, 347 (2013).
35. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* **173**, 53-61.e59 (2018).
36. L. Skov *et al.*, Detecting archaic introgression using an unadmixed outgroup. *PLOS Genetics* **14**, e1007641 (2018).
37. Y. Deng, R. Nielsen, Y. S. Song, Robust and accurate Bayesian inference of genome-wide genealogies for hundreds of genomes. *Nature Genetics* **57**, 2124-2135 (2025).
38. F. Baumdicker *et al.*, Efficient ancestry and mutation simulation with msprime 1.0. *Genetics* **220**, 19 (2021).
39. S. Burgess, A. Butterworth, S. G. Thompson, Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* **37**, 658-665 (2013).
40. J. Rozowsky *et al.*, The EN-TE resource of multi-tissue personal epigenomes & variant-impact models. *Cell* **186**, 1493-1511.e1440 (2023).
41. R. M. Baldarelli, C. L. Smith, M. Ringwald, J. E. Richardson, C. J. Bult, Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, iyae031 (2024).
42. J. Li *et al.*, Enhancer-promoter interaction maps provide insights into skeletal muscle-related traits in pig genome. *BMC Biology* **20**, 136 (2022).
43. Z. Pan *et al.*, Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nature Communications* **12**, 5848 (2021).
44. K. Karttunen *et al.*, Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nature Communications* **14**, 5313 (2023).
45. M. Goel, H. Sun, W.-B. Jiao, K. Schneeberger, SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology* **20**, 1-13 (2019).

46. G. Hickey *et al.*, Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology* **42**, 663-673 (2024).
47. E. L. Berdan *et al.*, Structural variants and speciation: multiple processes at play. *Cold Spring Harbor Perspectives in Biology* **16**, a041446 (2024).
- 5 48. J. Guo *et al.*, Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature Communications* **9**, 1865 (2018).
49. M. C. Turchin *et al.*, Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**, 1015-1019 (2012).
- 10 50. J.-H. Chen, X. Du, Z. Zheng, Dataset for “Ancient introgression drives wild boar expansion and phenotypic diversification of domestic pigs”. *ZENODO*, doi:10.5281/zenodo.15253420, (2025).
51. D. Wei *et al.*, New materials of early Pleistocene *Sus* from Sanhe Cave, Chongzuo, Guangxi, South China. *Acta Anthropologica Sinica* **32**, 63-76 (2013).
52. N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858 (2000).
- 15 53. H.-B. Xie *et al.*, African Suid Genomes Provide Insights into the Local Adaptation to Diverse African Environments. *Molecular Biology and Evolution* **39**, (2022).
54. W. Jiying, Y. Ying, F. Lixia, W. Huaizhong, Z. Qin, Genomic DNA extraction from hair sacs of pigs using modified phenol-chloroform method. *Hereditas (Beijing)* **32**, 752-756 (2010).
- 20 55. J. R. Arseneau, R. Steeves, M. Laflamme, Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Molecular ecology resources* **17**, 686-693 (2017).
56. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).
57. S. Koren *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**, 722-736 (2017).
- 25 58. B. J. Walker *et al.*, Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014).
59. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
- 30 60. O. Dudchenko *et al.*, De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95 (2017).
61. S. Wingett *et al.*, HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**, 1310 (2015).
62. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão, E. M. Zdobnov, BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution* **38**, 4647-4654 (2021).
- 35 63. J. Jurka *et al.*, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**, 462-467 (2005).
64. J. M. Flynn *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451-9457 (2020).
- 40 65. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
66. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018).
67. A. Shumate, B. Wong, G. Pertea, M. Pertea, Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS computational biology* **18**, e1009730 (2022).
68. J. Keilwagen, F. Hartung, M. Paulini, S. O. Twardziok, J. Grau, Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC bioinformatics* **19**, 189 (2018).
- 45 69. M. Stanke *et al.*, AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research* **34**, W435-W439 (2006).
70. B. L. Cantarel *et al.*, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* **18**, 188 (2008).
- 50 71. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
73. K. I. Kendig *et al.*, Sentieon DNaseq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. *Frontiers in Genetics* **10**, 736 (2019).
- 55 74. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
75. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

76. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880-1890 (2021).
77. J.-J. Jin *et al.*, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 1-31 (2020).
78. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
79. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780 (2013).
80. P. Librado, J. Rozas, DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452 (2009).
81. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
82. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599 (2007).
83. M. S. Peng *et al.*, DomeTree: a canonical toolkit for mitochondrial DNA analyses in domesticated animals. *Molecular Ecology Resources* **15**, 1238-1242 (2015).
84. B. M. Skinner *et al.*, The pig X and Y Chromosomes: structure, sequence, and evolution. *Genome Research* **26**, 130-139 (2016).
85. H. Ai *et al.*, Human-Mediated Admixture and Selection Shape the Diversity on the Modern Swine (*Sus scrofa*) Y Chromosomes. *Molecular Biology and Evolution* **38**, 5051-5065 (2021).
86. A. Criscuolo, A fast alignment-free bioinformatics procedure to infer accurate distance-based phylogenetic trees from genome assemblies. *Research Ideas and Outcomes* **5**, e36178 (2019).
87. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**, 559-575 (2007).
88. J. Felsenstein, PHYLIP-phylogeny inference package (version 3.2). *Cladistics* **5**, 164-166 (1989).
89. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
90. C. Zhang, S.-S. Dong, J.-Y. Xu, W.-M. He, T.-L. Yang, PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788 (2018).
91. R. R. Fitak, OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* **6**, bpab017 (2021).
92. A. H. Freedman *et al.*, Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLOS Genetics* **10**, e1004016 (2014).
93. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586-1591 (2007).
94. J. Gongora *et al.*, Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zoologica Scripta* **40**, 327-335 (2011).
95. Y. Wong *et al.*, A general and efficient representation of ancestral recombination graphs. *Genetics* **228**, iyae100 (2024).
96. J. Kelleher, A. M. Etheridge, G. McVean, Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS computational biology* **12**, e1004842 (2016).
97. O. Gascuel, BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14**, 685-695 (1997).
98. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496 (2011).
99. R. E. Handsaker *et al.*, Large multiallelic copy number variations in humans. *Nat Genet* **47**, 296-303 (2015).
100. J. P. Spence, Y. S. Song, Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances* **5**, eaaw9206 (2019).
101. F. Tortereau *et al.*, A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* **13**, 586 (2012).
102. N. V. Terekhanova, V. B. Seplyarskiy, R. A. Soldatov, G. A. Bazykin, Evolution of Local Mutation Rate and Its Determinants. *Molecular Biology and Evolution* **34**, 1100-1109 (2017).
103. C. Yan *et al.*, Inference of ancestral alleles in the pig reference genome. *Animal Genetics* **54**, 649-651 (2023).
104. G. Blom, *Statistical estimates and transformed beta-variables*. (Almqvist & Wiksell, Wiley/New York, 1958).

105. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821-824 (2012).
106. A. L. Price *et al.*, Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909 (2006).
107. S.-S. Dong *et al.*, LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics* **22**, bbaa227 (2021).
108. G. Hemani, K. Tilling, G. Davey Smith, Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genetics* **13**, e1007081 (2017).
109. G. Hemani *et al.*, The MR-Base platform supports systematic causal inference across the human phenome. *elife* **7**, e34408 (2018).
110. O. O. Yavorska, S. Burgess, MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology* **46**, 1734-1739 (2017).
111. M. Verbanck, MRPRESSO: Performs the Mendelian randomization pleiotropy residual sum and outlier (MR-PRESSO) test. *R package version 1*, (2017).
112. M. Verbanck, C.-Y. Chen, B. Neale, R. Do, Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature genetics* **50**, 693-698 (2018).
113. S. Burgess, S. G. Thompson, Interpreting findings from Mendelian randomization using the MR-Egger method. *European Journal of Epidemiology* **32**, 377-389 (2017).
114. J. M. Rees, A. M. Wood, F. Dudbridge, S. Burgess, Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLOS ONE* **14**, e0222362 (2019).
115. T. Ye, J. Shao, H. Kang, Debiased inverse-variance weighted estimator in two-sample summary-data Mendelian randomization. *The Annals of Statistics* **49**, 2079-2100 (2021).
116. A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
117. Y. Liu *et al.*, Trait correlated expression combined with eQTL and ASE analyses identified novel candidate genes affecting intramuscular fat. *BMC Genomics* **22**, 805 (2021).
118. G. H. Putri, S. Anders, P. T. Pyl, J. E. Pimanda, F. Zanini, Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943-2945 (2022).
119. J. Y. Hwang, E. B. Kim, H. Ka, C.-K. Lee, Identification of the Porcine XIST Gene and Its Differential CpG Methylation Status in Male and Female Pig Cells. *PLOS ONE* **8**, e73677 (2013).
120. J. Minks, W. P. Robinson, C. J. Brown, A skewed view of X chromosome inactivation. *The Journal of Clinical Investigation* **118**, 20-23 (2008).
121. G. Marçais *et al.*, MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* **14**, e1005944 (2018).
122. R. F. Balboa *et al.*, African bushpigs exhibit porous species boundaries and appeared in Madagascar concurrently with human arrival. *Nature Communications* **15**, 172 (2024).
123. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
124. D. Li *et al.*, Pangenome and genome variation analyses of pigs unveil genomic facets for their adaptation and agronomic characteristics. *iMeta* **3**, e257 (2024).
125. Y. Wu *et al.*, Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications* **9**, 918 (2018).
126. M. Gautier, A. Klassmann, R. Vitalis, rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources* **17**, 78-90 (2017).
127. K. Nidup, Genetic Structure and Biodiversity of Pigs (*Sus scrofa*) in South Asia and Papua New Guinea, Ph.D. Thesis, University of Sydney, (2011).

Acknowledgments

We are grateful to the editor and four anonymous reviewers for their insightful and constructive comments, which substantially improved the quality and clarity of this manuscript. We thank Dr. Brandon Stuart Gaut and Dr. J.J. Emerson from the University of California, Irvine, Dr. Manyuan Long (Department of Ecology and Evolution) and Dr. Bowei Kang (Public Health Sciences Department) from the University of Chicago, Dr. Yonggang Liu from Yunnan Agricultural University, Dr. Tao Wan from the Kunming Institute of Zoology, Chinese Academy of Sciences, and Mr. Xuefei He from West China Hospital for helpful discussions on the analyses. We thank Dr. Evgeniy Varisovich Kamaldinov and Dr. Valeriy Lavrentyevich Petukhov from University of Biotechnologies, Novosibirsk, Russia and Dr. Edward Okoth from ILRI for their invaluable help during the sampling process. We also thank the NKLCGI bioinformatics platform at Huazhong Agricultural University, BioHPC (DFG INST 86/2050-1 FUGG), and LRZ Munich for providing computing resources.

Funding: This study was supported by
NSFC innovation team (32221005 to SHZ);
NSFC Outstanding Youth (32125035 to XYL);
NSFC-CGIAR Cooperation (31961143020 to SHZ and JH);
Yazhouwan National Laboratory (2310SH01 and JL23YCKY01 to SHZ and JH);
IAEA CRP (D31030 to XD);
Modern Industrial Technology System of Pig (CARS-35 to SHZ);
Nepal NARC (LMBIS305/Yr2017 to NG);
GEF-UNEP-ILRI FAnGR RPSC GEF Project (1902 to LTT);
European Research Council (ERC-2013-StG-337574-UNDEAD or ERC-2019-StG-853272
PALAEOFARM or both to LAF, GL, DWGS);
Natural Environmental Research Council (NE/K005243/1 and NE/K003259/1 to LAF);
NERC-Ristekdikti Newton (NE/S007067/1 to LAF);
Taighde Éireann—Research Ireland (21/PATH-S/9515(T) to KD)
European Research Council Horizon 2020 (101220382-HERDPATH to KD);
ZIN RAS (125012800908-0 to MS);
Technological Innovation Research, Chengdu (2021-YF05-01331-SN to JHC);
R&D, West China Hospital (2020HXBH087 to JHC).

Author contributions:

Conceptualization: SHZ, JH, GL, LAF, JHC, XD.
Supervision: SHZ, JH, GL, LAF.
Resources: NG, SHZ, XYL, TNTT, ZY, PN, XLL, SM, SSH, MS, NS, LTT, AO, DP, MOF, AAB, GLLPS, AMS, MT, EVK, VLP, SS, SMFV, FH, HA, BDPS, RB, MC, XX.
Formal analysis: JHC, XD, ZZ, JZ, GH, CX, FJ, ZY, JL, ER, LLF, SLZ, PN.
Funding acquisition: SHZ, JH, XYL, XD, NG, LAF, GL, DWGS, KD, JHC, LTT, MS.
Writing – original draft: JHC.
Writing – review & editing: SHZ, JH, JHC, DWGS, YY, NV, GH, MH, XD, ZZ, IJ, KD, DF, MSW, JS, NG, CX, LZ, LAF, GL.

Competing interests:

All authors declare that they have no competing interests.

Data and materials availability:

Raw resequencing data generated by this study are archived under the National Center for Biotechnology Information (NCBI) BioProject accession numbers (PRJNA793285 and PRJNA851582). No new materials were generated for this study. The de novo genome assemblies, genomic annotations, important intermediate analysis files, significant P values and annotations for GWAS variants, QQ and Manhattan plots, and code/protocol are archived under Zenodo (50) and also <https://github.com/JianhaiChen/South-Asian-Wild-boar-and-pig>.

SUPPLEMENTARY MATERIALS

Materials and Methods

Supplementary Text

Figs. S1 to S43

Tables S1 to S25

References (51-117)

MDAR Reproducibility Checklist

Fig. 1. Population structure, geographic distribution, and genetic parameters. (A) Minimal topology with the highest posterior probability (0.896) inferred using AdmixtureBayes (15). The dashed line indicates admixture from *S. verrucosus* to ISEA wild boar (95% CI: 17.8–21.2%). Each node is labeled with its identifier and posterior probability. (B) PCA and geographic distribution of populations. Samples are colored by geographic region: South Asia (red), ISEA (light blue), MSEA (purple), East Asia (dark blue), and Europe (brown). (C) Genetic parameters (MAF > 0.05) across region groups: heterozygosity (H), nucleotide diversity (π), and inbreeding coefficient (F). Wilcoxon rank-sum test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, and ns: not significant). (D) Genome-wide frequencies of three alternative topologies inferred using Twisst (17) (5,000-SNPs windows): MSEA-sister (purple), South-Asia-sister (pink), and ISEA-sister (blue). "chrA" and "chrX" denote autosomes and the X chromosome. (E) Fine-scale recombination rates for genomic regions supporting the three topologies (Wilcoxon rank-sum test, **** $P < 0.0001$). (F) Dispersal routes of wild boar inferred from molecular divergence times estimated with MSMC2 (fig. S19).

Fig. 2. Phylogeography of X-linked haplotypes underlying body-size divergence. (A) X-chromosomal topologies of South Asian (SA) miniature and larger-bodied wild boar and domestic pigs. Left line plots show Twisst-derived topology weightings, color-coded to right-side topologies. (B) Phylogenetic trees for two X-chromosomal subregions (~44–57 and ~57–87 Mb) showing distinct placements of northern Chinese domestic pigs (100% bootstrap support). Node bars indicate Bayesian divergence times (median, 95% CI) calibrated by soft fossil constraints (33). The posterior age density for the miniature vs. larger-bodied clade split (~8.70–7.01 Mya) is shown. (C) Geographic distribution of miniature (purple) and larger-bodied (red) X-linked haplotypes. The larger-bodied haplotype includes South Asian (e.g., Nepalese Hurrah, Pakistani and Sri Lankan wild boar, plainland Bangladeshi domestic pigs), northern Asian/Chinese, and European populations. The “South Asian miniature” haplotype includes Nepalese wild boar and miniature breeds (e.g., Bampudke, eastern Chwanche) and southeastern Bangladeshi and northeastern Indian pigs. The “MSEA miniature” haplotype is found in wild boar and miniature domestic pigs from MSEA (e.g., Laotian, Vietnamese populations) and southern China (e.g., Wuzhishan, Luchuan, Diannan Small-Eared, Bama Miniature breeds). (D)

Hypothesized expansion of the larger-bodied X haplotype from South Asia to Europe and northern China.

Fig. 3. Association of the X-chromosomal introgression (~44–87 Mb) with body size and cold adaptation. (A) Manhattan plot for body size (Bonferroni $P < 2.04 \times 10^{-9}$) with the lead SNP labeled. (B) Manhattan plot for minimum December temperature ($P < 1.97 \times 10^{-9}$). (C) Causal effects between body size and cold adaptation using Mendelian randomization analyses. Points represent IVW fixed-effect estimated with 95% CI for significant associations (FDR < 0.05). Abbreviations: Precip, precipitation; Tmin, minimum temperature; LGM, Last Glacial Maximum; Bio10, mean temperature of warmest quarter; Bio19, precipitation of coldest quarter. (D) Schematic of the ASE-based PheWAS. RNA-Seq reads from F_1 offspring (large-bodied Duroc boars $\times$ miniature Luchuan sows) were aligned to parental haploid genomes to quantify ASE and then test body-size trait associations. (E) ASE-based PheWAS across the ~44–87 Mb region. Z-scores (β/SE) for genes (points) and distributions (boxplots) in large-bodied (purple) and miniature (blue) haplotypes (Bonferroni $P < 0.004$). (F) Enhancer enrichment heatmap (500-kb bins) across 14 tissues. Triangles represent multi-tissue regulatory segments (FDR < 0.05 , fold > 1 , and Z-score > 2.0 in at least 3 tissues). Color denotes tissue-normalized Z-scores of enhancer density (binomial test FDR in table S22). Top genes from (E) are shown above.

Fig. 4. Genetic differentiation and selection signals within the X-chromosomal introgressed region (~44–87 Mb). (A) Phased X-chromosomal haplotypes. Individual tracks are colored by haplotype group (blue: miniature; orange: large-bodied; $\Delta AF > 0.01$; 10-SNP windows; 10,000 randomly selected SNPs). Vertical dashed lines delineate introgression intervals at 44, 57, and 87 Mb. (B) F_{ST} analysis between miniature and large-bodied pigs. The brown line represents differentiation between South Asian miniature pigs (Nepalese Bampudke and Chwanche) and larger-bodied South Asian pigs. The cyan line represents differentiation between all miniature pigs and European large-bodied pigs. The orange line represents genetic differentiation between MSEA miniature pigs (Laotian Small-Eared, Vietnamese Mongcai and Ban, and Chinese Wuzhishan, Luchuan, Bama, and Diannan Small-Eared breeds) and larger-bodied pigs from northern China. (C–F) Selection summary statistics in domestic pigs within the introgressed X-chromosomal region: nucleotide diversity (π) (C), genetic differentiation from wild boar (F_{ST}) (D), Tajima's D (E), and allele frequency increase (ΔAF) (F) for body-size-increasing alleles relative to wild boar (alleles with GWAS $\beta > 0$ and $P < 2.04 \times 10^{-9}$).