Check for updates

PIG GENOMICS

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

Jian-Hai Chen[1,2,3]†‡, Xiaoyong Du[1,4,]†, Zhuqing Zheng[1,2]†, Neena Amatya Gorkhali[5]†, Changyi Xiao[4], Shilin Zhu[1], Jie Zhong[6], Liangliang Fu[1,2], Jingjin Li[1,2], Zhuofan Yan[4], Edward S. Ricemeyer[7,8], Fan Jiang[9], Thuy Nhien Thi Tran[1,2], Pan Ni[1], Nicholas VanKuren[3], Guanglin He[6], Yalan Yang[10], Supamit Mekchay[11], San San Hmwe[12], Mikhail Sablin[13], Nikica Šprem[14], Le Thi Thuy[15], Abdullah Omar[16], M. O. Faruque[17], Ali Akbar Bhuiyan[18], G. L. L. Pradeepa Silva[19], Amali Malshani Samaraweera[20], Debabrata Phukon[21], Mukesh Thakur[22], Saroj Sapkota[5], S. M. Farhad Vahidi[23,24], Faiz-ul Hussan[25], Hidayat Ashari[26,27], Bayu Dewantoro Putro Soewandi[1,27], Ivan Jakovlić[28], Kevin G. Daly[29,30], Ming-Shan Wang[31], Xiaolei Liu[1,2], Malavanh Chittavong[32], Daniel Falush[33], Jeffrey Spence[34,35], Richard Bishop[36], David W. G. Stanton[37,38], Xuewen Xu[1,2], Xinyun Li[1,2], Matthew W. Hahn[39,40], Lingzhao Fang[41], Laurent A. F. Frantz[7,37]*, Greger Larson[42]*, Jianlin Han[36,43,44]*, Shuhong Zhao[1,2,44]*

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 to 5 million years ago) in South Asia from an archaic Suinae lineage that introduced a ~43 megabase 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 that was later shaped by natural and artificial selection.

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 [~0.5 to 1 million years ago (Ma)] (*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 megabase (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 and 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 and S3 and tables S5 to S7). This model of South Asian wild boar as the 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 (PP) = 0.896; Fig. 1A], whereas alternative models involving interspecific introgression into South Asian wild boar received negligible support (PP = 0.0001; fig. S5). Principal components analysis (PCA) further revealed a clear genetic differentiation of South Asian wild boar and domestic pigs from other groups (Fig. 1B, fig. S6, and table S1).

The South Asian group exhibited higher historical effective population size (Ne), greater genetic diversity, faster linkage disequilibrium (LD) decay, and a lower inbreeding coefficient (Fig. 1C, figs. S7 to S9, and table S8). Of the ~104.78 million single-nucleotide polymorphisms (SNPs) genotyped in our dataset, 45.47% (~47.64 million) were previously unreported in the dbSNP database (*16*). Of these new SNPs, 70.29% were common [minor allele frequency (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 and table S10). This inference was robust to local nonmonophyly: Even among nonmonophyletic windows, most of the windows placed South Asian samples as 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 *Sus 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 thousand years ago (ka) (303 to 341 ka) (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 ($A_{\mathrm{Sver-ISEA}}$) (figs. S15 to S17 and tables S12 and S13). After controlling for GC content, CpG density, local mutation-rate proxy, and mappability, the correlation between the recombination rate and $A_{\mathrm{Sver-ISEA}}$ remained moderate to strong on autosomes ($\rho = 0.637$ to 0.983) and the X chromosome ($\rho = 0.562$ to 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

Downloaded from https://www.science.org on September 26, 2026

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% of 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. Multiple sequentially Markovian coalescent 2 (MSMC2) analysis (*27*, *28*) estimated the South Asian wild boar separation from other lineages at 658 to 695 ka, substantially predating the divergence among all pairs of non–South Asian populations (113 to 360 ka; 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 pig 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 and tables S14 and S15), with most admixture events (92.3%, $n = 13$) occurring in the past ~73 generations. By 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 and 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 to 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 to 7.01 Ma (Fig. 2B and fig. S22).

We further tested whether the large-body–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- to 87-Mb region (Sprime scores > 95th percentile; HMMix, PP > 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- to 87-Mb region than for the autosomal background (fig. S25), consistent with acquisition of the large-body–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 to 57 and ~57 to 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 to 3.05 Ma. 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 in *S. scrofa* (Fig. 2, C and D).

Association between the X-chromosomal introgression and body size

Marked divergence in the introgressed region (~44 to 87 Mb) between miniature and large-bodied pigs suggests an association with body size. To test this hypothesis, we performed a genome-wide association

Downloaded from https://www.science.org on September 26, 2026

[1]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan, China. [2]Frontiers Science Center for Animal Breeding and Sustainable Production, Wuhan, China. [3]Department of Ecology and Evolution, The University of Chicago, Chicago, IL, USA. [4]Key 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, China. [5]National Animal Breeding and Genetics Research Centre, Nepal Agricultural Research Council, Khumaltar, Nepal. [6]Institute of Rare Diseases, West China Hospital of Sichuan University, Sichuan University, Chengdu, China. [7]Chair of Animal Systems Genomics, Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany. [8]Bond Life Sciences Center, University of Missouri, Columbia, MO, USA. [9]Institute of Animal Sciences and Veterinary Medicine, Hubei Academy of Agricultural Sciences, Key Laboratory of Animal Embryo Engineering and Molecular Breeding of Hubei Province, Wuhan, China. [10]Department of Human Genetics, The University of Chicago, Chicago, IL, USA. [11]Department of Animal and Aquatic Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai, Thailand. [12]Dagon University, Yangon, Myanmar. [13]Zoological Institute of the Russian Academy of Sciences, St. Petersburg, Russia. [14]Department of Fisheries, Apiculture, Wildlife Management and Special Zoology, University of Zagreb Faculty of Agriculture, Zagreb, Croatia. [15]National Institute of Animal Sciences, Hanoi, Vietnam. [16]National Engineering Laboratory for Animal Breeding, Key Lab of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, China Agricultural University, Beijing, China. [17]Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh. [18]Livestock Division, Bangladesh Agricultural Research Council (BARC), Dhaka, Bangladesh. [19]Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka. [20]Department of Animal Science, Faculty of Animal Science and Export Agriculture, Uva Wellassa University, Badulla, Sri Lanka. [21]Animal Husbandry and Veterinary Department, Government of Assam, Guwahati, India. [22]Mammal and Osteology Section, Zoological Survey of India, Kolkata, India. [23]Animal Science Research Department, Qom Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Qom, Iran. [24]Department of Systems and Synthetic Biology, Agricultural Biotechnology Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran. [25]Department of Breeding and Genetics, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Punjab, Pakistan. [26]National Research and Innovation Agency (BRIN), Cibinong, West Jawa, Indonesia. [27]Research Organization for Agriculture and Food, National Research and Innovation Agency (BRIN), Cibinong, West Jawa, Indonesia. [28]State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, and College of Ecology, Lanzhou University, Lanzhou, China. [29]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [30]School of Agriculture and Food Science, University College Dublin, Dublin, Ireland. [31]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. [32]Department of Livestock and Fisheries, Faculty of Agriculture, National University of Laos, Vientiane, Lao People's Democratic Republic. [33]Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China. [34]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. [35]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. [36]International Livestock Research Institute (ILRI), Nairobi, Kenya. [37]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [38]Cardiff School of Biosciences, Cardiff University, Cardiff, UK. [39]Department of Biology, Indiana University, Bloomington, IN, USA. [40]Department of Computer Science, Indiana University, Bloomington, IN, USA. [41]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark. [42]PalaeoBARN, School of Archaeology, University of Oxford, Oxford, UK. [43]CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China. [44]Yazhouwan National Laboratory, Sanya, Hainan, China. ***Corresponding author. Email: laurent.frantz@palaeo.vetmed.uni-muenchen.de (L.A.F.F.); greger.larson@arch.ox.ac.uk (G.L.); hanjianlin@yzwlab.cn (J.H.); zhaoshuhong@yzwlab.cn (S.Z.)** †These authors contributed equally to this work. ‡Present address: University of California, Irvine, Irvine, CA, USA.

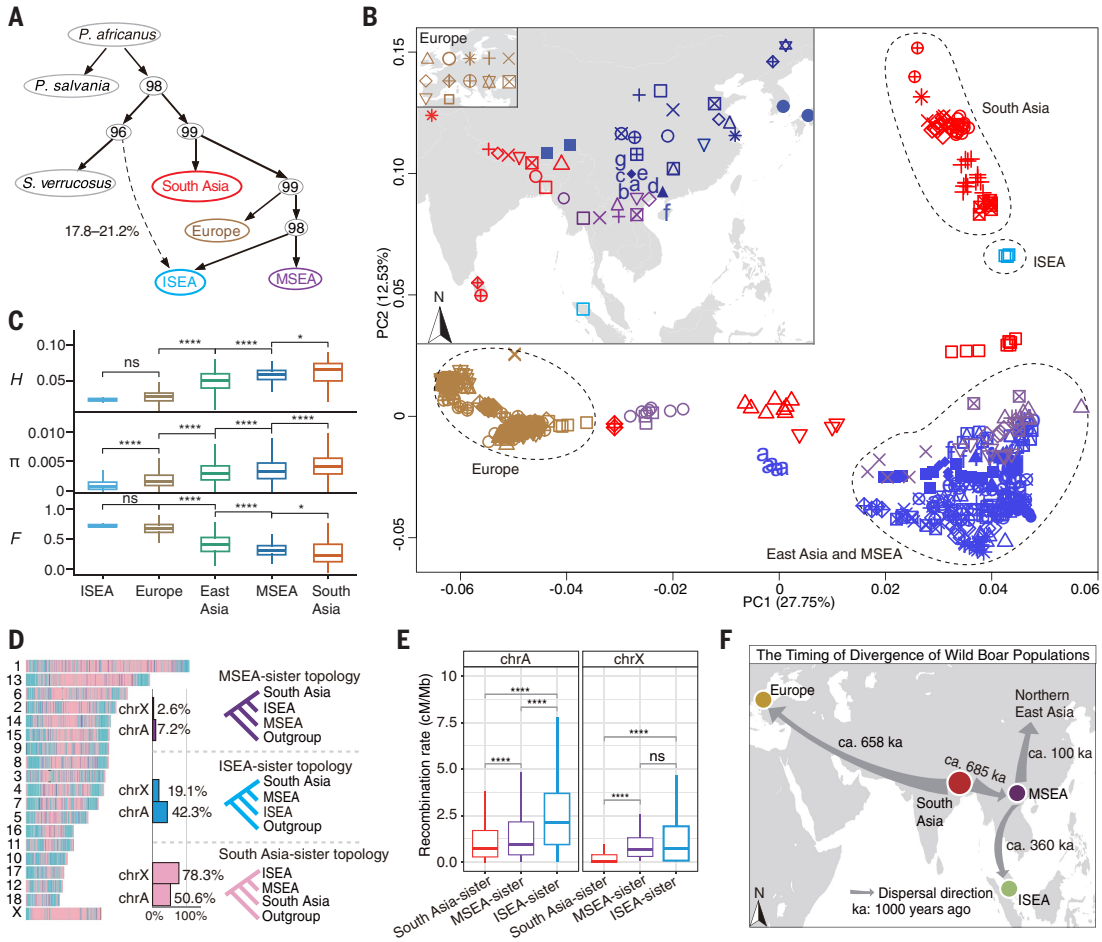


Fig. 1. Population structure, geographic distribution, and genetic parameters. (**A**) Minimal topology with the highest PP (0.896) inferred using AdmixtureBayes (*15*). The dashed line indicates admixture from *S. verrucosus* to ISEA wild boar [95% confidence interval (CI) = 17.8 to 21.2%]. Each node is labeled with its PP. (**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 (*$P < 0.05$; ****$P < 0.0001$; and ns, not significant). (**D**) Genome-wide frequencies of three alternative topologies inferred using Twisst (*17*) (5000 SNP 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).

study (GWAS) for breed-level body size using a linear mixed model with an autosomal relatedness matrix, sex, and principal components (PCs) to account for population structure ($\lambda_{autosome} = 1.010$; Fig. 3A, fig. S27, and table S16). Nearly all significant SNPs mapped to the ~44- to 87-Mb region (16,365 out of 16,383; 99.9%), with the strongest association signal concentrated in the first subregion (~44 to 57 Mb; Bonferroni $P < 2.04 \times 10^{-9}$). LD analysis revealed uniformly high D' (mean 0.963) but heterogeneous squared correlation coefficients (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 (~10^2-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 ka) and recent decades (1979–2013) ($\lambda = 0.992$ to 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 in the ~44- to 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 [inverse-variance weighted (IVW) method β: −0.21 to −0.11, false discovery rate (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 to 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 F_1 resource population generated from large-bodied Duroc sires × miniature Luchuan dams ($n = 189$, 77 females; Fig. 3D and table S19) and analyzed the association between ASE of muscle-expressed genes

Downloaded from https://www.science.org on September 26, 2026

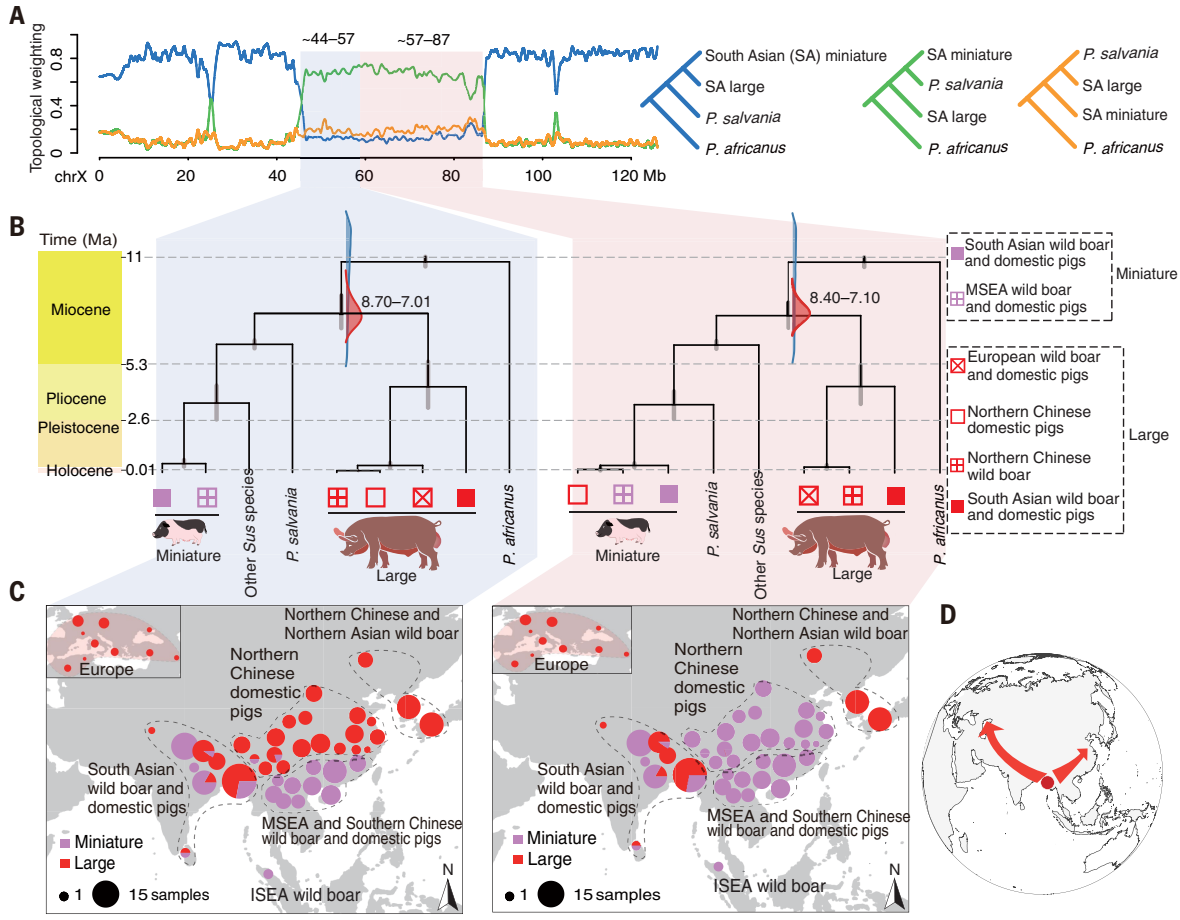


Fig. 2. Phylogeography of X-linked haplotypes underlying body-size divergence. (**A**) X-chromosomal topologies of South Asian (SA) miniature and larger-bodied wild boar and domestic pigs. Left line plots show Twisst-derived topology weightings, color-coded to right-side topologies. (**B**) Phylogenetic trees for two X-chromosomal subregions (~44 to 57 and ~57 to 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 versus larger-bodied clade split (~8.70 to 7.01 Ma) 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 Chinese and Northern Asian, and European populations. The "South Asian miniature" haplotype includes Nepalese wild boar and miniature breeds (e.g., Bampudke and 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 and Vietnamese populations) and southern China (e.g., Wuzhishan, Luchuan, Diannan Small-Eared, and Bama Miniature breeds). (**D**) Hypothesized expansion of the larger-bodied X haplotype from South Asia to Europe and northern China.

and carcass straight length, a representative adult body-size trait. Genome-wide comparison showed that this correlation peaked in the ~44- to 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 F_1 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 and 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-body–haplotype-biased expression correlated positively with body-size traits (e.g., *ZC4H2*, *EFNB1*, and *MAGED1*), whereas 68% (23/34) involving miniature–haplotype-biased expression showed negative correlations (e.g., *AMER1*, *SMC1A*, and *GLA*). Several of these genes are annotated for the 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- to 87-Mb region across 14 major pig tissues using integrative chromatin-profiling datasets, including chromatin immunoprecipitation sequencing for histone modifications and assay for transposase-accessible chromatin with sequencing (ATAC-seq) (*42*, *43*). We identified six enhancer-enriched candidate regulatory subregions (0.5 Mb in width; 44.0 to 44.5, 56.0 to 56.5, 57.5 to 58.0, 63.0 to 63.5, 83.0 to 83.5, and 84.5 to 85.0 Mb) with moderate to strong enhancer enrichment (Fig. 3F, fig. S34, and 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 to 58.0 Mb). These chromatin annotations

Downloaded from https://www.science.org on September 26, 2026

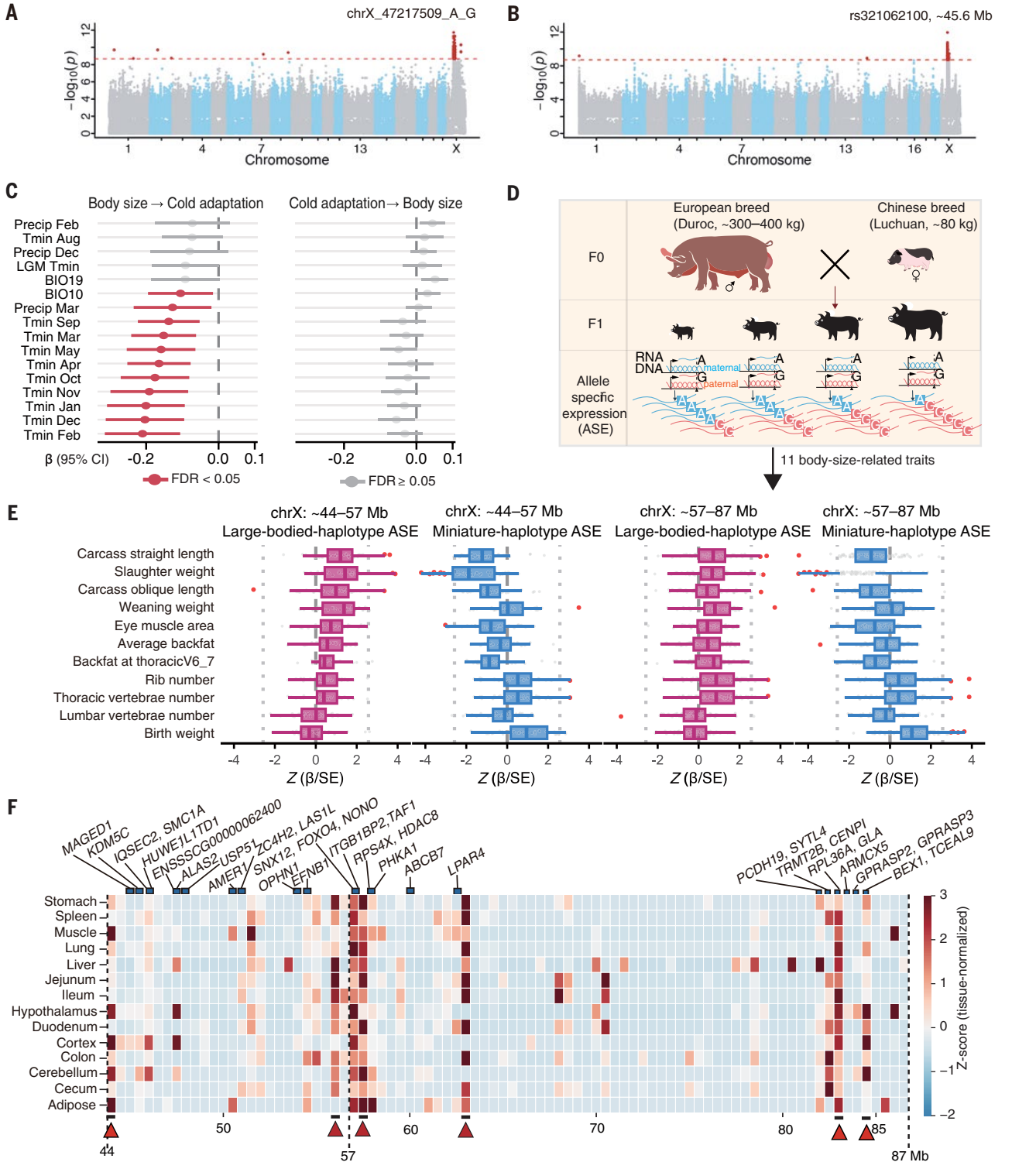


Fig. 3. Association of the X-chromosomal introgression (~44 to 87 Mb) with body size and cold adaptation. (**A**) Manhattan plot for body size (Bonferroni $P < 2.04 \times 10^{-9}$) with the lead SNP labeled. (**B**) Manhattan plot for minimum December temperature ($P < 1.97 \times 10^{-9}$). (**C**) Causal effects between body size and cold adaptation using MR analyses. Points represent IVW fixed-effect estimates with 95% CI for significant associations (FDR < 0.05). BIO10, mean temperature of warmest quarter; BIO19, precipitation of coldest quarter; LGM, Last Glacial Maximum; Precip, precipitation; Tmin, minimum temperature. (**D**) Schematic of the ASE-based PheWAS. RNA-sequencing reads from F_1 offspring (large-bodied Duroc boars × 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- to 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 multitissue regulatory segments (FDR < 0.05, fold > 1, and *Z* score > 2.0 in at least three tissues). Color denotes tissue-normalized *Z* scores of enhancer density (binomial test FDR in table S22). Top genes from (E) are shown above.

Downloaded from https://www.science.org on September 26, 2026

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- to 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 and 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 Synteny and Rearrangement Identifier (SyRI) (*45*) and validated using a pangenome graph–based approach (*46*). On the X chromosome, we detected 134 medium-to-large SVs (100 bp to 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 and tables S24 and S25). Most of these SVs (126/134, 94.0%) were concentrated within the ~44- to 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 in the X-chromosomal introgressed haplotype in domestic pig 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 pig populations, likely favoring increased body size (*9*), acting on the limited standing genetic variation in the introgressed haplotype.

To further test this hypothesis, we computed several summary statistics, including fixation index (F_{ST}), Tajima's *D*, allele frequency change (ΔAF), and extended haplotype homozygosity (EHH) decay. In the ~44- to 57-Mb subregion, F_{ST} was 3.0- to 3.9-fold higher than the genome-wide background across northern Chinese, European, and South Asian pigs (top 0.55 to 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 rank sum 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. Before this dispersal, admixture (~3 to 5 Ma) 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

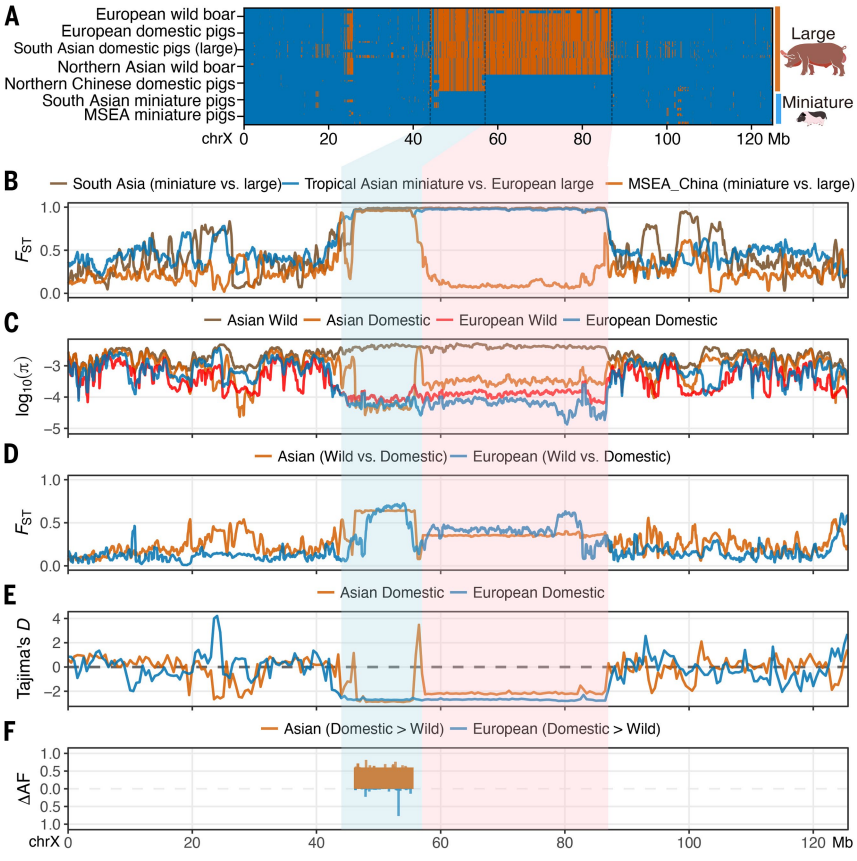


Fig. 4. Genetic differentiation and selection signals in the X-chromosomal introgressed region (~44 to 87 Mb). (**A**) Phased X-chromosomal haplotypes. Individual tracks are colored by haplotype group (blue, miniature; orange, large-bodied; ΔAF > 0.01; 10-SNP windows; 10,000 randomly selected SNPs). 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** to **F**) Selection summary statistics in domestic pigs in 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 $P < 2.04 \times 10^{-9}$).

Downloaded from https://www.science.org on September 26, 2026

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 SVs accumulated in 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.

REFERENCES AND NOTES

1. G. Larson *et al.*, *Science* **307**, 1618–1621 (2005).
2. J. Chen *et al.*, *Ecol. Evol.* **8**, 8102–8114 (2018).
3. M. F. Rothschild, A. Ruvinsky, Eds., *The Genetics of the Pig* (CABI, 2011).
4. M. A. M. Groenen *et al.*, *Nature* **491**, 393–398 (2012).
5. L. Frantz *et al.*, *Annu. Rev. Anim. Biosci.* **4**, 61–85 (2016).
6. L. A. F. Frantz *et al.*, *Genome Biol.* **14**, R107 (2013).
7. H. Ai *et al.*, *Nat. Genet.* **47**, 217–225 (2015).
8. L. A. F. Frantz *et al.*, *Nat. Genet.* **47**, 1141–1148 (2015).
9. C. Lega, P. Raia, L. Rook, D. Fulgione, *Holocene* **26**, 327–332 (2016).
10. J. K. Lunney *et al.*, *Sci. Transl. Med.* **13**, eabd5758 (2021).
11. K. Nidup, D. D. Joshi, J. Gongora, C. Moran, *Biodiversity (Nepean)* **11**, 26–33 (2010).
12. H. M. Briggs, *Modern Breeds of Livestock* (Macmillan, 1969).
13. U. Albarella, K. Dobney, P. Rowley-Conwy, *Environ. Archaeol.* **14**, 103–136 (2009).
14. R. Maier *et al.*, *eLife* **12**, e85492 (2023).
15. S. V. Nielsen *et al.*, *PLOS Genet.* **19**, e1010410 (2023).
16. S. T. Sherry *et al.*, *Nucleic Acids Res.* **29**, 308–311 (2001).
17. S. H. Martin, S. M. Van Belleghem, *Genetics* **206**, 429–438 (2017).
18. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
19. J. K. Pickrell, J. K. Pritchard, *PLOS Genet.* **8**, e1002967 (2012).
20. N. B. Edelman *et al.*, *Science* **366**, 594–599 (2019).
21. M. Malinsky, M. Matschiner, H. Svardal, *Mol. Ecol. Resour.* **21**, 584–595 (2021).
22. L. A. Frantz, O. Madsen, H. J. Megens, M. A. Groenen, K. Lohse, *Mol. Ecol.* **23**, 5566–5574 (2014).
23. J. Kamm, J. Terhorst, R. Durbin, Y. S. Song, *J. Am. Stat. Assoc.* **115**, 1472–1487 (2020).
24. M. Schumer *et al.*, *Science* **360**, 656–660 (2018).
25. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, *PLOS Biol.* **17**, e2006288 (2019).
26. J. B. Pease, M. W. Hahn, *Evolution* **67**, 2376–2384 (2013).
27. M. Zhang, Q. Yang, H. Ai, L. Huang, *Genomics Proteomics Bioinformatics* **20**, 1040–1052 (2022).
28. S. Schiffels, R. Durbin, *Nat. Genet.* **46**, 919–925 (2014).
29. M. Bosse *et al.*, *Nat. Commun.* **5**, 4392 (2014).
30. B. Yang *et al.*, *Genet. Sel. Evol.* **49**, 71 (2017).
31. G. Hellenthal *et al.*, *Science* **343**, 747–751 (2014).
32. D. J. Lawson, G. Hellenthal, S. Myers, D. Falush, *PLOS Genet.* **8**, e1002453 (2012).
33. L. Liu *et al.*, *Nat. Commun.* **10**, 1992 (2019).
34. N. Zamani *et al.*, *BMC Genomics* **14**, 347 (2013).
35. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, *Cell* **173**, 53–61.e9 (2018).
36. L. Skov *et al.*, *PLOS Genet.* **14**, e1007641 (2018).
37. Y. Deng, R. Nielsen, Y. S. Song, *Nat. Genet.* **57**, 2124–2135 (2025).
38. F. Baumdicker *et al.*, *Genetics* **220**, iyab229 (2022).
39. S. Burgess, A. Butterworth, S. G. Thompson, *Genet. Epidemiol.* **37**, 658–665 (2013).
40. J. Rozowsky *et al.*, *Cell* **186**, 1493–1511.e40 (2023).
41. R. M. Baldarelli *et al.*, *Genetics* **227**, iyae031 (2024).
42. J. Li *et al.*, *BMC Biol.* **20**, 136 (2022).
43. Z. Pan *et al.*, *Nat. Commun.* **12**, 5848 (2021).
44. K. Karttunen *et al.*, *Nat. Commun.* **14**, 5313 (2023).
45. M. Goel, H. Sun, W.-B. Jiao, K. Schneeberger, *Genome Biol.* **20**, 277 (2019).
46. G. Hickey *et al.*, *Nat. Biotechnol.* **42**, 663–673 (2024).
47. E. L. Berdan *et al.*, *Cold Spring Harb. Perspect. Biol.* **16**, a041446 (2024).
48. J. Guo *et al.*, *Nat. Commun.* **9**, 1865 (2018).
49. M. C. Turchin *et al.*, *Nat. Genet.* **44**, 1015–1019 (2012).
50. J.-H. Chen, X. Du, Z. Zheng, Dataset for "Ancient introgression drives wild boar expansion and phenotypic diversification of domestic pigs," Zenodo (2025); https://doi.org/10.5281/zenodo.15253420.

ACKNOWLEDGMENTS

We thank four anonymous reviewers for their insightful and constructive comments, which substantially improved the quality and clarity of this manuscript. We also thank B. S. Gaut and J. J. Emerson from the University of California, Irvine; M. Long (Department of Ecology and Evolution) and B. Kang (Public Health Sciences Department) from the University of Chicago; Y. Liu from Yunnan Agricultural University; T. Wan from the Kunming Institute of Zoology, Chinese Academy of Sciences; and X. He from West China Hospital for helpful discussions on the analyses. We thank E. V. Kamaldinov and V. L. Petukhov from the University of Biotechnologies, Novosibirsk, Russia, and E. 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 Shu.Z.); NSFC Outstanding Youth (32125035 to Xin.L.); NSFC-CGIAR Cooperation (31961143020 to Shu.Z. and J.H.); Yazhouwan National Laboratory (2310SH01 and JL23YCKY01 to Shu.Z. and J.H.); IAEA CRP (D31030 to X.D.); Modern Industrial Technology System of Pig (CARS-35 to Shu.Z.); Nepal NARC (LMBIS305/Yr2017 to N.A.G.); GEF-UNEP-ILRI FAnGR RPSC GEF Project (1902 to L.T.T.); European Research Council (ERC-2013-StG-337574-UNDEAD or ERC-2019-StG-853272 PALAEOFARM or both to L.A.F.F., G.L., D.W.G.S.); Natural Environmental Research Council (NE/K005243/1 and NE/K003259/1 to L.A.F.F.); NERC-Ristekdikti Newton (NE/S007067/1 to L.A.F.F.); Taighde Éireann—Research Ireland (21/PATH-S/9515(T) to K.G.D.); European Research Council Horizon 2020 (101220382-HERDPATH to K.G.D.); ZIN RAS (125012800908-0 to M.S.); Technological Innovation Research, Chengdu (2021-YF05-01331-SN to J.-H.C.); and R&D, West China Hospital (2020HXBH087 to J.-H.C.). **Author contributions:** Conceptualization: Shu.Z., J.H., G.L., L.A.F.F., J.-H.C., X.D.; Supervision: Shu.Z., J.H., G.L., L.A.F.F.; Resources: N.A.G., Shu.Z., Xin.L., T.N.T.T., Z.Y., P.N., Xia.L., S.M., S.S.H., M.S., N.S., L.T.T., A.O., D.P., M.O.F., A.A.B., G.L.L.P.S., A.M.S., M.T., S.S., S.M.F.V., F.H., H.A., B.D.P.S., R.B., M.C., X.X.; Formal analysis: J.-H.C., X.D., Z.Z., J.Z., G.H., C.X., F.J., Z.Y., J.L., E.S.R., Lia.F., Shi.Z., P.N.; Funding acquisition: Shu.Z., J.H., Xin.L., X.D., N.A.G., L.A.F.F., G.L., D.W.G.S., K.G.D., J.-H.C., L.T.T., M.S.; Writing—original draft: J.-H.C.; Writing—review & editing: Shu.Z., J.H., J.-H.C., D.W.G.S., Y.Y., N.V., G.H., M.W.H., X.D., Z.Z., I.J., K.G.D., D.F., M.-S.W., J.S., N.A.G., C.X., Lin.F., L.A.F.F., G.L. **Competing interests:** All authors declare that they have no competing interests. **Data, code, 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 remaining frozen tissue specimens collected for this study, together with the extracted DNA and RNA samples, are available from Shu.Z. subject to a material transfer agreement (MTA). 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. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adq7553
Materials and Methods; Supplementary Text; Figs. S1 to S43; Tables S1 to S25; References (*51–127*); MDAR Reproducibility Checklist

Submitted 29 May 2024; resubmitted 8 August 2025; accepted 5 August 2026

10.1126/science.adq7553

Downloaded from https://www.science.org on September 26, 2026

Science
AAAS

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

Jian-Hai Chen, Xiaoyong Du, Zhuqing Zheng, Neena Amatya Gorkhali, Changyi Xiao, Shilin Zhu, Jie Zhong, Liangliang Fu, Jingjin Li, Zhuofan Yan, Edward S. Ricemeyer, Fan Jiang, Thuy Nhien Thi Tran, Pan Ni, Nicholas VanKuren, Guanglin He, Yalan Yang, Supamit Mekchay, San San Hmwe, Mikhail Sablin, Nikica Šprem, Le Thi Thuy, Abdullah Omar, M. O. Faruque, Ali Akbar Bhuiyan, G. L. L. Pradeepa Silva, Amali Malshani Samaraweera, Debabrata Phukon, Mukesh Thakur, Saroj Sapkota, S. M. Farhad Vahidi, Faiz-ul Hussan, Hidayat Ashari, Bayu Dewantoro Putro Soewandi, Ivan Jakovli#, Kevin G. Daly, Ming-Shan Wang, Xiaolei Liu, Malavanh Chittavong, Daniel Falush, Jeffrey Spence, Richard Bishop, David W. G. Stanton, Xuewen Xu, Xinyun Li, Matthew W. Hahn, Lingzhao Fang, Laurent A. F. Frantz, Greger Larson, Jianlin Han, and Shuhong Zhao

Science **393** (6818), . DOI: 10.1126/science.adq7553

Editor's summary

Archeological evidence has suggested an origin in tropical Asia for wild boar, but genetic sampling from modern pig populations in this region has been limited. Chen *et al*. examined 745 genomes from pigs and wild boar from 58 populations across mainland and island Southeast Asia. They found that South Asian wild boar are deeply diverged from European wild boar and domesticated pigs, although gene flow between these groups has been extensive. This extended sampling revealed that a haplotype on the X chromosome associated with larger body size was likely introgressed from an extinct lineage. Allele-specific expression in crosses between miniature and large pig breeds suggested potential genes underlying this important agronomic trait. —Corinne Simonti

View the article online
https://www.science.org/doi/10.1126/science.adq7553
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from https://www.science.org on September 26, 2026