Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes

Laurent A F Frantz[1,2], Joshua G Schraiber[3,4], Ole Madsen[1], Hendrik-Jan Megens[1], Alex Cagan[5], Mirte Bosse[1], Yogesh Paudel[1], Richard P M A Crooijmans[1], Greger Larson[2] & Martien A M Groenen[1]

Traditionally, the process of domestication is assumed to be initiated by humans, involve few individuals and rely on reproductive isolation between wild and domestic forms. We analyzed pig domestication using over 100 genome sequences and tested whether pig domestication followed a traditional linear model or a more complex, reticulate model. We found that the assumptions of traditional models, such as reproductive isolation and strong domestication bottlenecks, are incompatible with the genetic data. In addition, our results show that, despite gene flow, the genomes of domestic pigs have strong signatures of selection at loci that affect behavior and morphology. We argue that recurrent selection for domestic traits likely counteracted the homogenizing effect of gene flow from wild boars and created 'islands of domestication' in the genome. Our results have major ramifications for the understanding of animal domestication and suggest that future studies should employ models that do not assume reproductive isolation.

The rise of agriculture, which occurred approximately 10,000 years ago, was one of the most important transitions in human history[1]. During the Neolithic Revolution, the domestication of plant and animal species led to a major shift in subsistence, from a hunter-gatherer to a sedentary agricultural lifestyle, which ultimately resulted in the development of complex societies. The process of animal domestication led to striking morphological and behavioral changes in domesticated individuals in comparison to their wild progenitors[2]. Traditionally, this process has often been viewed as being directed by humans and involving strong bottlenecks in the domestic population (corresponding to founder events due to the selection of only a few individuals at the beginning of domestication) and reproductive isolation between wild and domestic forms[3–5]. This straightforward model provides an attractive theoretical framework for geneticists because key events such as the geographical origin and timeframe of domestication are well defined. The assumption of reproductive isolation eases the interpretation of genetic data from domestic and wild forms. For instance, under this model, geneticists have interpreted phylogenetic affinities of domestic animals with multiple, geographically divergent wild populations as evidence of frequent, independent domestication origins in multiple species[6–11].

However, this view conflicts with zooarchaeological evidence that shows that domestication episodes are rare and that domesticated forms diffuse out from a limited number of core regions[12–14]. Moreover, there is a growing body of empirical and theoretical archaeological work[12,15,16] that challenges the simplicity of traditional models. In the new, more complex models, prehistoric domestication of animals is viewed as mainly having been unintentional[12,15,16], and neither reproductive isolation nor strong intentional selection are viewed as having been as crucial and widespread as previously thought. Instead, domestication is seen as a long-term, diffuse process[17], involving gene flow (during as well as after domestication) between wild and domestic populations[18] and with emphases on multiple, taxon-specific human-animal relationships[15,16]. The possibility of post-domestication gene flow between domestic animals and their wild progenitors and a lack of strong domestication bottlenecks[18] are key predictions from this new framework that contrast with more traditional models of domestication. Moreover, extensive gene flow between wild and domestic populations violates the assumptions of traditional models of domestication and has major ramifications for studies that attempt to infer the spatial and chronological origin of domestication using genetic data[12].

Here we focus on pig domestication using genome-wide data sets of modern domestic pigs and wild boars. Pigs were domesticated independently in Anatolia[19] and the Mekong valley about 9,000 years before the present[20]. Furthermore, analyses of ancient mitochondrial DNA (mtDNA) found that the first domestic pigs in Europe were transported by early farmers from the Near East into Europe

[1]Animal Breeding and Genomics Group, Wageningen University, Wageningen, the Netherlands. [2]Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK. [3]Department of Integrative Biology, University of California, Berkeley, Berkeley, California, USA. [4]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [5]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Correspondence should be addressed to L.A.F.F. (laurent.frantz@arch.ox.ac.uk).

Received 21 November 2014; accepted 10 August 2015; published online 31 August 2015; doi:10.1038/ng.3394

around 7,500 years before the present, concordant with zooarchaeological evidence for a single domestication origin of western Eurasian domestic pigs[21]. However, a few thousand years after their introduction, domestic pigs in Europe had completely lost the Near Eastern mtDNA signatures and instead acquired mtDNA haplotypes found in local European wild boar[21,22]. These findings suggest that early domestic populations experienced post-domestication gene flow from wild boar populations that were not involved in the Anatolian domestication process[12,21] and that European wild boar was not independently domesticated.

Further mtDNA analyses of ancient Anatolian material demonstrated that, by 2,500 years before the present, local mtDNA haplotypes were replaced by European haplotypes. This result suggests extensive mobile swine herding throughout Europe and Anatolia[22], consistent with both archaeological and historical evidence, as well as limited management and selection up until the Industrial Revolution in the nineteenth century[23,24]. Thus, under a complex model of domestication, mtDNA replacement in ancient European and Anatolian pigs was the result of post-domestication gene flow, loose pig management and mobile swine herding. We expect such phenomena to have left a strong signal of gene flow from wild boars in the genome of modern domestic pigs.

However, although unsupported by any zooarchaeological evidence, the observed mtDNA turnovers could also be interpreted as a *de novo* domestication of a population of European wild boars rather than the result of post-domestication gene flow from wild boars. Moreover, because of mode of inheritance and limited resolution, small mtDNA markers provide a very limited impression of gene flow, making it impossible to test these hypotheses. Thus, the hypothesis of complex domestication in pigs has yet to be tested with the resolution and confidence afforded by the large-scale analysis of nuclear markers. In addition, unlike with horses and donkeys, intentional interbreeding between pigs and wild boars confers no clear advantage to pig production and is thought to have occurred mainly unintentionally[18]. Lastly, there is a clear morphological and behavioral dichotomy between wild boars and domestic pigs that is evident in modern animals as well as in the zooarchaeological record[25–28]. Thus, the possibility of unintentional gene flow between wild boars and domestic pigs also raises questions regarding the mechanisms behind the maintenance of traits that differentiate domestic and wild forms[18].

Here we fit models of domestication to a genome-wide data set from over 100 wild and domestic pigs. Our main aim is to test whether models following a traditional, linear explanation or those involving a more complex, reticulation process better fit the modern data. More generally, we assess whether the zooarchaeological evidence for a single geographically restricted domestication of (Western) pigs in Anatolia[12,13,21] is compatible with the assumption of a traditional model of domestication involving reproductive isolation and strong bottlenecks.

RESULTS

We evaluated the support of multiple models for the domestication of pigs in Europe and Asia. Our analysis focused on 103 genomes from European wild boars (EUW)[6] and European commercial and non-commercial (rare or historical) domestic breeds (EUD)[29] (**Supplementary Table 1**). In addition, the data set comprised multiple populations of Asian wild boar (ASW) and Asian domestic pigs (ASD; **Supplementary Table 1**). We also included a Javan Warty pig (*Sus verrucosus*) as an outgroup (**Supplementary Fig. 1**). To better understand the early process of domestication, we sampled a range of wild boar populations, from Asia and all major European Pleistocene refugia[6], non-commercial and commercial European pig breeds[29] and Asian pig breeds. To test key predictions of the complex domestication framework described above, we fit simple but informative models to the genomic data set using approximate Bayesian computation (ABC; Online Methods).

Gene flow between wild boars and domestic pigs

We first tested the hypothesis of gene flow between wild boars and domestic pigs. More precisely, we asked whether reproductive isolation of wild boars and domestic pigs was compatible with the zooarchaeological evidence that pigs were domesticated only twice, independently in Anatolia and China. To address this, we first evaluated the fit of the traditional model in which domestication is modeled as two parallel events in Asia and Europe. In this model, domestication takes place at time t_1 in Europe and time t_2 in Asia and involves no gene flow between wild and domestic forms (reproductive isolation) or between domestic pigs from Asia and Europe (**Fig. 1**, top). We then compared this null model to five models involving different patterns of continuous mixture, including gene flow (i) within wild boar populations, (ii) within domestic pig populations, (iii) between wild and domestic forms, etc. (**Supplementary Fig. 2**). By comparing these six models, we found that a model involving gene flow between domestic and wild forms (within Asia and Europe) as well as between different domestic populations (between Europe and Asia) received substantial support (posterior probability (PP) = 0.93) when compared to any other model tested in this study (**Fig. 1**, middle). Moreover, we found that all models including gene flow in this study provided an extreme improvement of fit in comparison to the null model with no gene flow (**Supplementary Fig. 3** and **Supplementary Table 2**). Thus, our explicit modeling framework provides very strong evidence that reproductive isolation between wild and domestic forms was not maintained during and after domestication in Asia and Europe. However, our current data set does not allow us to conclude the extent of admixture in different breeds. Thus, differential admixture, from multiple divergent populations of wild boar, could have contributed to the complex population structure observed among pig breeds (**Figs. 2** and **3**).

To further assess these findings, we evaluated the ancestry of wild boars and domestic pigs, using our genome sequences as well as a data set of over 600 pigs (from the same populations as in the genome-wide data) that were genotyped on the PorcineSNP60 array (**Supplementary Note**). Our analyses showed that EUD and ASD were both paraphyletic, whereas EUW was monophyletic (**Fig. 2b**, **Supplementary Fig. 4** and **Supplementary Note**). Moreover, in our principal-component analysis (PCA), the first components discriminated between domestic and wild forms in Asia (**Fig. 2c**) and in Europe (**Fig. 2d**). In addition, we found that EUD and ASD displayed more substructure than EUW and ASW (**Fig. 2**) and shared a large amount of ancestry with wild populations (**Fig. 3**). The paraphyly, substructure and ancestry of EUD and ASD are difficult to reconcile with the assumptions of a linear, spatially restricted model of domestication. Instead, our findings provide further evidence of a complex domestication process that involved gene flow between wild boars and domestic pigs. Moreover, we found that gene flow between wild boars and domestic pigs in Europe was strongly asymmetrical (the migration rate from EUW to EUD was much higher than the rate from EUD to EUW; **Supplementary Fig. 5** and **Supplementary Note**).

Lastly, we also found evidence that Asian and European domestic pigs exchanged genetic material. This finding is consistent with previous studies and is most likely the result of European importation of Chinese pigs during the nineteenth century to improve European

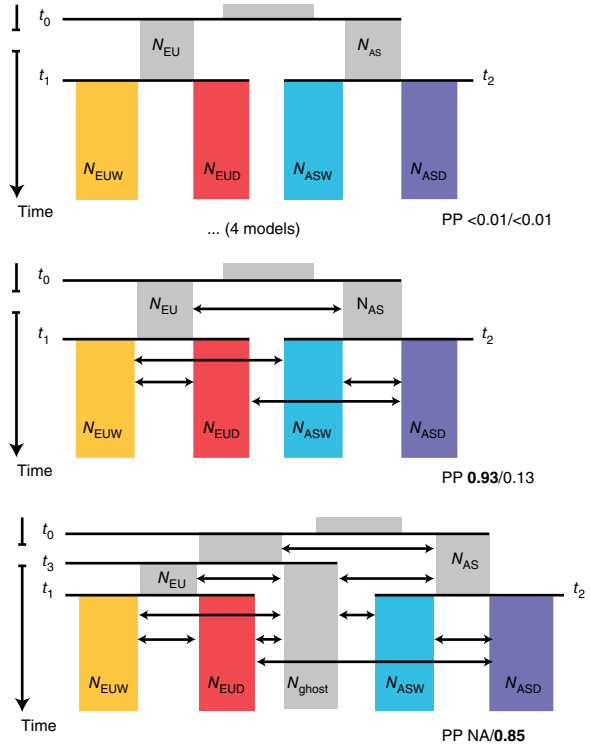


Figure 1 Schematics of the models tested in this study. The model-testing approach compared six or seven models. Double-headed arrows represent migrations that were modeled as two independent, continuous parameters. Two of the six models not including the ghost population are displayed: one without gene flow (null model; top) and one with gene flow between wild boars (ASW and EUW) and domestic pigs (ASD and EUD) as well as between domestic pig populations and between wild boar populations (full model; middle); the four additional models tested in this study are displayed in **Supplementary Figure 2**. A seventh model comprised the full model with gene flow between the wild and domestic populations and a ghost population (bottom). In the posterior probability notation, the probabilities appearing first were computed without the ghost model (comparing six models in total) and the probabilities appearing second were computed with seven models in total, including the ghost model (bottom). Bold values indicate the most likely model for each comparison (with and without the ghost model). EU, European; AS, Asian; NA, not applicable.

commercial breeds[24,30–32]. However, gene flow between the domestic populations (EUD and ASD) was very limited relative to gene flow between wild and domestic populations (between EUW and EUD and between ASW and ASD; **Supplementary Fig. 5** and **Supplementary Note**). This result is not surprising given our sampling strategy that focused on non-commercial European domestic breeds that are less likely to be admixed with Asian domestic pigs[24,29]. We conclude that this small amount of gene flow between domestic pig populations suggests that admixture between European and Asian domestic pig populations had little influence on our conclusion that gene flow occurred between wild boars and domestic pigs on both sides of Eurasia (**Supplementary Note**).

Together, these findings demonstrate that domestic pigs do not constitute a homogeneous genetic group, as would be expected under a simple model involving human-driven domestication. Instead, domestic pig breeds are a genetic mosaic of different wild boar populations. Thus, modern genetic data from domestic pigs can only be reconciled with zooarchaeological evidence for a restricted domestication process if modeled with continuous gene flow between wild and domestic populations.

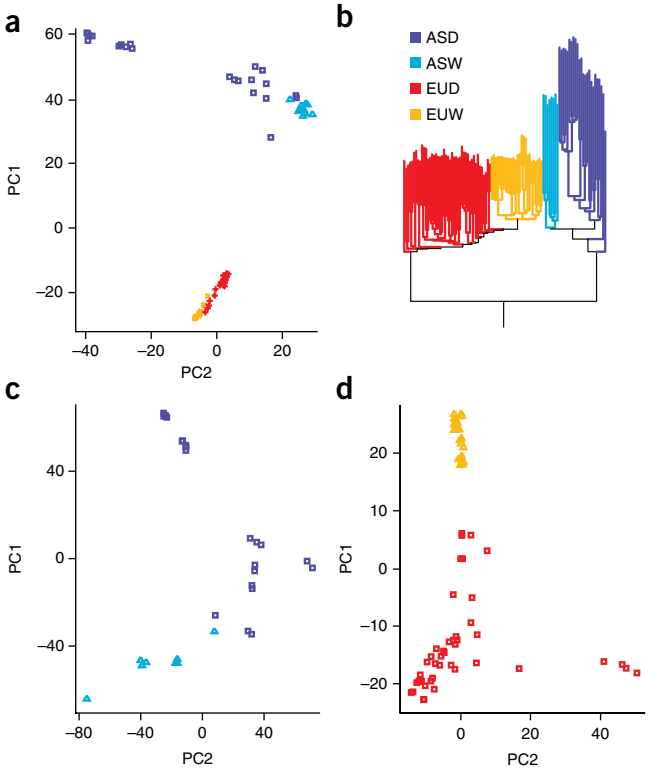


Figure 2 Ancestry of wild boars and domestic pigs in Europe and Asia. (**a**) Results of PCA using 500,000 randomly chosen SNPs from our genome data set. (**b**) Neighbor-joining tree, based on an identity-by-state (IBS) distance matrix (Online Methods), depicting the relationship between the genome sequences used (except for the outgroup) in this study. (**c**) PCA for Asian samples using 500,000 randomly chosen SNPs ascertained in Asian pigs. (**d**) PCA for European samples using 500,000 randomly chosen SNPs ascertained in European pigs. PC1, principal component 1; PC2, principal component 2.

No evidence of a domestication bottleneck

We also tested for the presence of a strong population bottleneck associated with domestication. To do so, we estimated the posterior distribution of demographic parameters using 10,000 simulations retained out of 2,000,000. Under the assumption of a simple linear model of domestication with no gene flow and strong, intentional selection by humans, we would expect a strong bottleneck in domestic populations. Overall, our results are consistent with a population decline in both EUW and EUD (**Fig. 4**). These findings support previous results, as well as historical evidence, demonstrating that Pleistocene glaciations resulted in long-term population decline in European wild boars[30,32,33]. However, this population decline was more pronounced in EUW than in EUD (**Fig. 4**). In addition, we found that the effective population size of EUD (Ne_{EUD} = 20,563, 95% highest prior density interval (HPDI) = 3,724–73,907) was more than twice the effective population size of EUW (Ne_{EUW} = 8,497, 95% HPDI = 1,555–33,873). However, the wide 95% HPDIs make these results difficult to interpret by themselves (**Supplementary Table 3** and **Supplementary Note**). Nevertheless, we found that the Ne_{EUD} parameter was greater than Ne_{EUW} in 66% of the 10,000 retained simulations (77% when retaining the 1,000 simulations closest to the observed values and 85% when

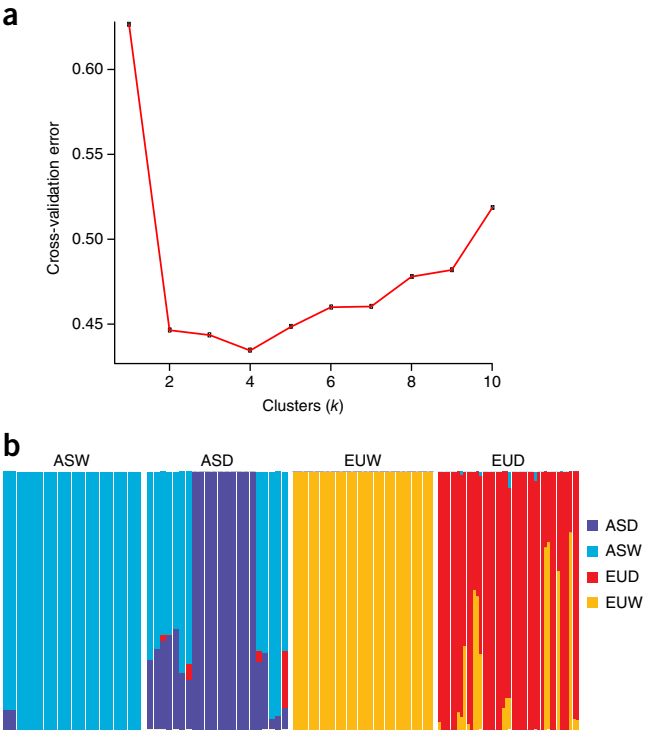


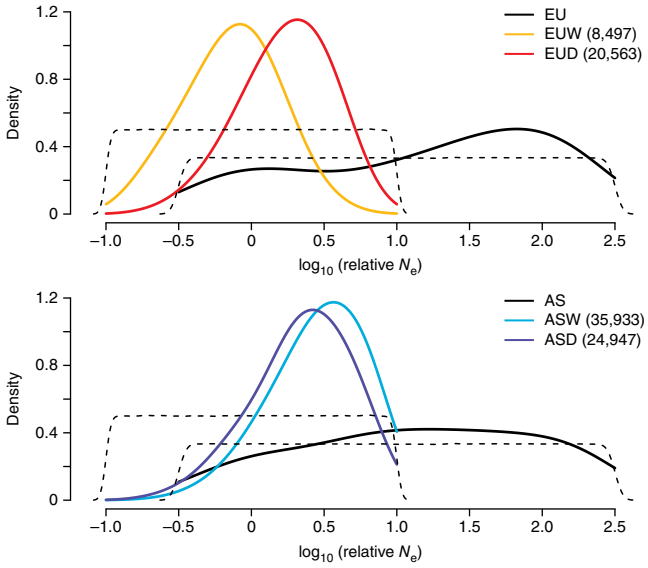
Figure 3 Model-based ancestry analysis. ADMIXTURE[74] analysis was performed to estimate the optimum number of clusters (k) in the data set using the same 500,000 randomly chosen SNPs as in PCA. (**a**) Cross-validation errors for diverse k values. As shown, $k = 4$ minimizes the cross-validation error. (**b**) Ancestry of each sample using $k = 4$ clusters.

retaining 100 simulations). Moreover, our low estimate of *Ne* in EUW corroborates previous findings that patterns of heterozygosity (such as long runs of homozygosity) are consistent with recent inbreeding in EUW[34]. This low *Ne* value is most likely due to a series of strong bottlenecks in the wild European populations, caused by overhunting and loss of suitable habitat[24,30,34,35].

Together, these results do not support the existence of a strong domestication bottleneck in European domestic pigs and instead support the contention that continuous gene flow from multiple genetically and geographically distinct wild boar populations likely inflated the effective population size of EUD. In other words, gene flow from EUW into EUD most likely increased the apparent substructure within this population (**Figs. 2** and **3**, and **Supplementary Figs. 3**, **6** and **7**) and probably affected our current *Ne* estimates. Moreover, the effective population size of EUW may well have been much higher at the time of domestication. Thus, our results do not completely rule out a domestication bottleneck; however, the most parsimonious explanation of the data does not support the existence of one. In the case of ASW and ASD, our current sampling does not allow us to draw strong conclusions because of the highly heterogeneous ancestry in ASW and the possibility of admixture with divergent species[33,36] (**Fig. 4** and **Supplementary Note**).

Gene flow from a second population of wild boars

We showed that a model incorporating continuous gene flow between wild boars and domestic pigs was significantly (PP > 0.99; **Supplementary Table 2**) more compatible with zooarchaeological evidence than a traditional hypothesis assuming reproductive isolation. Despite this fact, we only modeled gene flow from a population of wild boars that we assumed to be closely related to the population constituting the source of domestication. However, the high degree of substructure observed in EUD (**Fig. 2d**) suggests that EUW is unlikely to be the sole source of the genetic variation found in European domestic populations. We therefore tested the hypothesis that another population of wild boars that is either extinct (owing to hunting pressure and habitat loss) or was not sampled in our analysis (for example, wild populations from central Eurasia, where we did not sample) might have also contributed to the gene pool of modern European domestic pigs. For this testing, we used a model that was similar to our best-fitting model (**Fig. 1**, middle) but had an additional 'ghost' population that split from the EUW-EUD branch during the Pleistocene (**Fig. 1**, bottom) and acts as a step between ASW and EUW (migration ASW ↔ ghost population ↔ EUW but also ghost population ↔ EUD). This model provided a substantial improvement in fit and received high posterior support (PP = 0.85; **Fig. 1**, bottom, and **Supplementary Table 2**). This result shows that the mobile herding of domestic pigs across Europe most likely resulted in gene flow from at least one wild boar population that was genetically divergent from the population involved in the domestication process in Anatolia.



Positive selection counteracted the effect of gene flow

Our analysis demonstrates that gene flow between wild and domestic forms was a ubiquitous feature of domestication and post-domestication processes in pigs. Thus, extensive gene flow from wild boars into domestic pigs during and after domestication raises questions regarding the mechanisms behind the maintenance of the clear morphological and behavioral differences observed between domestic pigs and wild boars. Intentional or unintentional, artificial selection could have counteracted the effect of gene flow and resulted in morphological and behavioral differentiation between wild boars and domestic pigs.

To assess the importance of selection in the genome of domestic pigs in the face of gene flow, we conducted a scan for positive selection

Figure 4 Posterior density distributions of demographic parameters. Population sizes are shown as the relative population proportion (the ratio of the current population size to the population size before time 0 (t_0); **Fig. 1**). Each value in parentheses corresponds to the mode of the effective population size (N_e). The dashed lines represent the prior distributions.

using SweeD[37,38]. SweeD computes the composite likelihood ratio (CLR) of a selective sweep model over a neutral model. Such a test can be very sensitive to demography and migration[39]. To correct for these types of effects, we generated an expected cumulative distribution function (CDF) of CLR by simulating 10,000 large genomic fragments (3 Mb) under our best-fitting model, without the ghost population (**Fig. 1**, middle). The parameters for these simulations were drawn from the posterior distributions (see the **Supplementary Note** for details). We used this CDF to compute the *P* values for all empirical CLR values in the genome (Online Methods and **Supplementary Note**). We identified 1,953 and 1,014 10-kb regions (out of the 214,007 and 216,062 regions considered) with $P < 0.01$ in the genomes of European and Asian domestic pigs, respectively (**Supplementary Table 4**), providing conservative identification of positively selected segments in the genomes of wild boars and domestic pigs. We also investigated the interaction between artificial selection and purifying selection. To do so, we correlated derived allele frequency (DAF) in each population and F_{ST} (fixation index) between wild and domestic populations with conservation scores downloaded from the UCSC Genome Browser (**Supplementary Note**). We found a significant negative correlation between conservation scores and F_{ST} or DAF (**Supplementary Note**). These results suggest that derived alleles occurring at high frequency are less conserved than expected by chance and that highly conserved sites are less affected by artificial selection (as indicated by the lower F_{ST} values).

We then examined swept regions (indicating positive artificial selection) private to each population (**Supplementary Note**). The swept regions in domestic pigs (EUD and ASD) contained genes that were significantly enriched ($P < 0.05$) for Gene Ontology (GO) terms related to multiple developmental processes for bones, teeth and the nervous system (**Supplementary Tables 5** and **6**). These terms comprised multiple gene candidates related to height in pigs[40,41] and cattle[42,43] (*PLAG1*, *NCPAG*, *PENK*, *RPS20* and *LYN* in EUD and *LEMD3* and *UKP1B* in ASD) and nervous system development and maintenance (*NRTN*, *SEMA3C*, *PLXNC1*, *AAK1*, *RAB35* and *FRS2*)[44–55] and genes directly influencing behavior (for example, aggressiveness and feeding; *APBA2*, *MC4R*, *RCAN1* and *BAIAP3*)[56–63] (**Supplementary Tables 7** and **8**). These results suggest that domestication and/or post-domestication selection for behavioral and morphological traits was important in Asian and European domestic pigs and most likely counteracted the effect of continuous gene flow in certain parts of the genome.

However, the mechanism behind this maintenance remains unknown. One possibility is that there was recurrent selection for similar traits. This phenomenon might have resulted in parallel

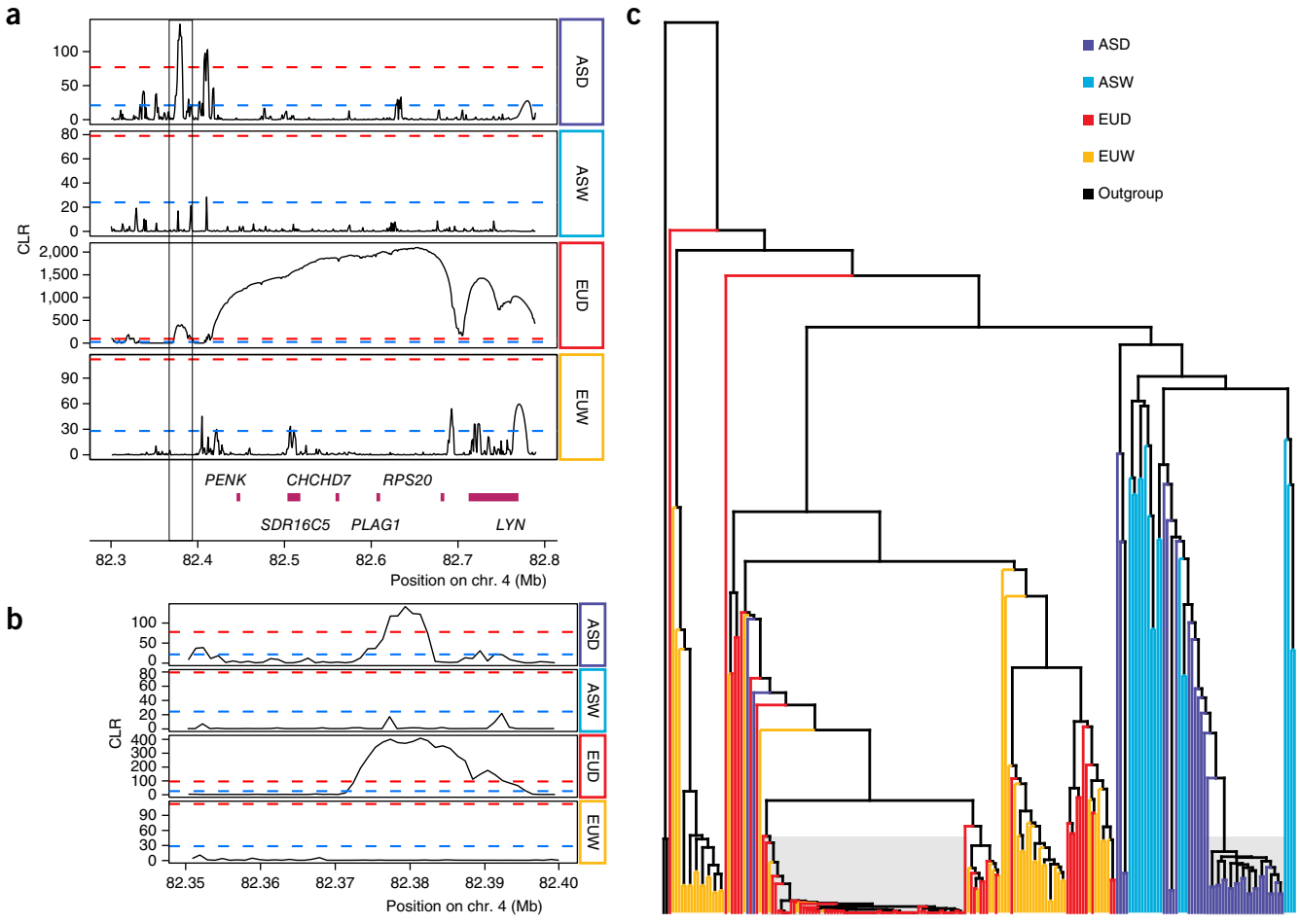


Figure 5 Example of a parallel sweep in ASD and EUD. (**a**) CLR values in the *PLAG1* region. Dashed blue and red lines represent *P*-value thresholds of 0.05 and 0.01, respectively. The boxed area indicates the position of the parallel swept region in **b**. (**b**) CLR values in the parallel swept region a few thousand basepairs upstream of the *PLAG1* region. (**c**) Genealogy of phased haplotypes (**Supplementary Note**) for the swept region in **b**. The shaded areas highlight the very short branch lengths that are the result of a selective sweep. The shaded area on the left (Europe) contains 64 haplotypes from EUD (>72% of the total EUD haplotypes) and 2 haplotypes from EUW (<4% of the total EUW haplotypes). The shaded area on the right (Asia) contains 24 haplotypes from ASD (>54% of the total ASD haplotypes) and no ASW haplotypes.

sweeps at the same loci in both ASD and EUD. To investigate this possibility, we looked for signals of parallel selection for the two independent domestication episodes (ASD and EUD). We identified regions in the genome with a significant CLR value ($P < 0.01$) in both ASD and EUD but not in ASW and EUW (**Supplementary Table 4**).

To rule out admixture between ASD and EUD as the cause for observing an overlapping significant signal, we conducted a phylogenetic analysis in each region separately (**Supplementary Note**). The genealogies of some of these regions showed a signal consistent with introgression between EUD and ASD (for example, see **Supplementary Fig. 8**). However, we found one region of particular interest that seems to have been swept independently in EUD and ASD (**Fig. 5**). Interestingly, although this swept region does not overlap any gene, the region is just a few kilobases upstream of the region with the highest CLR value in EUD (**Fig. 5a**). We validated the presence of this swept region using various haplotype[64] and frequency-based methods (**Supplementary Figs. 9** and **10**, and **Supplementary Note**). This region (with the highest CLR in EUD) has been shown to have a strong effect on body height and stature in pigs[40,41] and cattle[42,43]. In particular, variation in this region explains up to 18% of the difference in body length between wild boars and commercial EUD breeds[40]. Given the importance of this region for morphology in commercial EUD pigs[41], it is possible that human-mediated selection for similar traits in Asian and European domestic pigs resulted in parallel sweeps at the same locus. Parallel selection of this form may be responsible for some of the morphological convergence within as well as between the two independent domestication episodes in Europe and Asia. Thus, although the phenotypic effect of this swept region is still unclear, this region provides a particularly interesting candidate to further study the possibility of convergence between ASD and EUD.

DISCUSSION

The generation of larger amounts of genomic data with ever greater resolution is allowing the complexity of the domestication process to be more fully appreciated. The commensurate advancements in theoretical and empirical perspectives are allowing more sophisticated models to be tested and a greater understanding of animal domestication. In this study, we demonstrate that the assumptions of traditional models, such as reproductive isolation and strong domestication bottlenecks, are incompatible with the zooarchaeological evidence of a geographically restricted domestication process in pigs. Instead, our model-testing approach shows that continuous gene flow from wild boars to domestic pigs is necessary to reconcile modern genetic data with the zooarchaeological evidence. Moreover, we demonstrate that, in western Eurasia, gene flow most likely involved at least a second unsampled (possibly even extinct) population of wild boars that was quite divergent from the source of domestication. Gene flow from this population is most likely the result of mobile domestic swine herding, as predicted by zooarchaeologists[18,22]. The unequivocal evidence of widespread gene flow presented here should stimulate future studies to test this hypothesis in other taxa, as it was suggested in dogs[65] and horses[66]. It will be interesting in the future to test the hypothesis that the appearance and maintenance of reproductive isolation is correlated with the biological specificity (behavior) as well as mode of domestication (direct, as in rabbits, versus commensal, as in pigs and dogs[15]) in different domestic species.

Extensive gene flow from wild boars raises questions regarding the maintenance of morphological and behavioral traits in domestic pigs. Our study shows extensive evidence of selection at candidate genes that influence anatomical and nervous system development, suggesting that selection may have counteracted the homogenizing effect of gene flow and maintained the genetic basis for the morphological and behavioral dichotomy observed between wild boars and domestic pigs. In addition, our results show that regions close to genes governing morphological traits have been subject to selection in parallel in Asia and Europe. In the context of speciation, studies focusing on systems in which gene flow is common have identified genomic regions that show excessive interspecies divergence[67]. These studies have suggested that such regions may be resistant to gene flow and likely allowed for the maintenance of species-specific characteristics (constituting genomic 'islands of speciation'). Here we hypothesize that an analogous process took place during pig domestication. By recurrently selecting for similar traits through artificial selection but allowing for gene flow, farmers have created genomic 'islands of domestication', which we define as genomic regions governing domestic traits that are thus less affected by gene flow from wild boars. However, it is unclear whether these swept regions involved recurrent selection of different haplotypes from standing genetic variation in wild boars or are the result of selection from *de novo* mutations. Thus, our results highlight a list of candidate genes that will provide future studies with the means to further test these hypotheses.

In this study, we only discuss artificial selection in the context of its role in counteracting the effect of gene flow. Nevertheless, artificial selection may also have favored the retention of alleles that were introgressed from wild forms. Such a phenomenon was proposed for loci associated with night blindness in horse[68] and immunity in diverse domestic mammals[69]. However, such a hypothesis remains difficult to assess in a broad context given the complex genetic architectures underlying domestic traits[70]. Moreover, assessing adaptive introgression in domestic pigs is very challenging because of the poor representation of ancestral genetic diversity in modern wild boars. Nevertheless, given the wide geographical range occupied by the wild ancestor of domestic pigs[6] as well as the many closely related taxa capable of interbreeding[33,36], future studies may unravel examples of adaptive backcrossing between wild and domestic pigs.

Lastly, it is important to underline the limitations of modern DNA sequences and traditional domestication models in determining the origin and time of domestication for animals, as well as in identifying the genes involved during domestication. Indeed, extensive gene flow clearly violates the assumptions of traditional models and likely eroded most of the signal that could be used to infer time and geographical parameters[71–73]. It is therefore important to apply caution when conducting comparative analyses of modern genetic material from wild and domestic animals. However, future sequencing of ancient DNA, together with more realistic modeling frameworks, such as the one presented here, will provide the necessary information not only to determine the geographical origin and time of domestication for animals and plants but also to identify genes involved during domestication and will ultimately substantially enhance knowledge of this fascinating evolutionary process.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Raw reads have been deposited in the European Nucleotide Archive (ENA) under accession PRJEB9922.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank B. Peter for his help and guidance during the model-fitting step of the analysis as well as for kindly sharing his code. We are also indebted to D. Wegmann for providing us with the latest version of ABCtoolbox. We also thank K. Lohse for his insights during the conception of the project and D. Caspani for his help with the figures. We thank B. Dibbits for his help during the laboratory work. This project is financially supported by the European Research Council under the European Community's 256 Seventh Framework Programme (FP7/2007–2013)/ERC (SelSweep), grant agreements 249894 and ERC-2013-StG 337574-UNDEAD. J.G.S. was supported by US National Institutes of Health grant R01-GM40282 and National Science Foundation postdoctoral fellowship DBI-1402120.

AUTHOR CONTRIBUTIONS

M.A.M.G., L.A.F.F. and G.L. designed the study. R.P.M.A.C. provided the samples. L.A.F.F., H.-J.M., O.M., M.B. and Y.P. aligned and filtered the data. L.A.F.F. and J.G.S. performed the modeling. L.A.F.F., J.G.S. and A.C. performed the selection scan. L.A.F.F., J.G.S., G.L. and M.A.M.G. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Childe, V.G. *The Dawn of European Civilization* 6th edn. (Routledge & Kegan Paul, 1968).
2. Darwin, C. *The Variation of Animals and Plants Under Domestication* (John Murray, 1868).
3. Price, E.O. *Animal Domestication and Behavior* (CABI Publishing, 2002).
4. Driscoll, C.A., Macdonald, D.W. & O'Brien, S.J. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. USA* **106**, 9971–9978 (2009).
5. O'Connor, T.P. Wild or domestic? Biometric variation in the cat *Felis silvestris* Schreber. *Int. J. Osteoarchaeol.* **17**, 581–595 (2007).
6. Larson, G. *et al.* Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618–1621 (2005).
7. Hanotte, O. *et al.* African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336–339 (2002).
8. Luikart, G. *et al.* Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. USA* **98**, 5927–5932 (2001).
9. Naderi, S. *et al.* The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. USA* **105**, 17659–17664 (2008).
10. Pedrosa, S. *et al.* Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc. Biol. Sci.* **272**, 2211–2217 (2005).
11. Vilà, C. *et al.* Widespread origins of domestic horse lineages. *Science* **291**, 474–477 (2001).
12. Larson, G. & Fuller, D.Q. The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
13. Zeder, M.A. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. *Proc. Natl. Acad. Sci. USA* **105**, 11597–11604 (University of Oklahoma Press, 2008).
14. Macneish, R.S. *The Origins of Agriculture and Settled Life* (University of Oklahoma Press, 1992).
15. Zeder, M.A. in *Harlan II: Biodiversity in Agriculture: Domestication, Evolution and Sustainability* (eds. Damania, A. & Gepts, P.) 227–229 (Univ. California Press, 2011).
16. Vigne, J.-D. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *C. R. Biol.* **334**, 171–181 (2011).
17. Dobney, K. & Larson, G. Genetics and animal domestication: new windows on an elusive process. *J. Zool. (Lond.)* **269**, 261–271 (2006).
18. Marshall, F.B., Dobney, K., Denham, T. & Capriles, J.M. Evaluating the roles of directed breeding and gene flow in animal domestication. *Proc. Natl. Acad. Sci. USA* **111**, 6153–6158 (2014).
19. Ervynck, A., Hongo, H., Dobney, K. & Meadow, R. Born free? New evidence for the status of *Sus scrofa* at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient* **27**, 47–73 (2001).
20. Cucchi, T., Hulme-Beaman, A., Yuan, J. & Dobney, K. Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. *J. Archaeol. Sci.* **38**, 11–22 (2011).
21. Larson, G. *et al.* Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc. Natl. Acad. Sci. USA* **104**, 15276–15281 (2007).
22. Ottoni, C. *et al.* Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol. Biol. Evol.* **30**, 824–832 (2013).
23. Albarella, U., Manconi, F. & Trentacoste, A. in *Ethnozooarchaeology. The Present and Past of Human-Animal Relationships* (eds. Albarella, U. & Trentacoste, A.) 143–159 (Oxbow Books, 2011).
24. White, S. From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. *Environ. Hist.* **16**, 94–120 (2011).
25. Evin, A. *et al.* The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* **40**, 735–743 (2013).
26. Albarella, U., Davis, S.J.M., Detry, C.P. & Rowley-Conwy, P. Pigs of the 'Far West': the biometry of *Sus* from archaeological sites in Portugal. *Anthropozoologica* **40**, 27–54 (2005).
27. Rowley-Conwy, P., Albarella, U. & Dobney, K. Distinguishing wild boar from domestic pigs in prehistory: a review of approaches and recent results. *J. World Prehist.* **25**, 1–44 (2012).
28. Owen, J. *et al.* The zooarchaeological application of quantifying cranial shape differences in wild boar and domestic pigs (*Sus scrofa*) using 3D geometric morphometrics. *J. Archaeol. Sci.* **43**, 159–167 (2014).
29. Porter, V. *Pigs: A Handbook to the Breeds of the World* (Helm Information, 1993).
30. Groenen, M.A.M. *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012).
31. Bosse, M. *et al.* Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392 (2014).
32. Bosse, M. *et al.* Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol. Ecol.* **23**, 4089–4102 (2014).
33. Frantz, L.A. *et al.* Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol.* **14**, R107 (2013).
34. Bosse, M. *et al.* Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* **8**, e1003100 (2012).
35. Frantz, L.A.F., Madsen, O., Megens, H.-J., Groenen, M.A.M. & Lohse, K. Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island Southeast Asian *Sus* species during the Plio-Pleistocene climatic fluctuations. *Mol. Ecol.* **23**, 5566–5574 (2014).
36. Ai, H. *et al.* Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
37. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**, 1566–1575 (2005).
38. Pavlidis, P., Živkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
39. Huber, C.D., Nordborg, M., Hermisson, J. & Hellmann, I. Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana*. *Mol. Biol. Evol.* **31**, 3026–3039 (2014).
40. Andersson-Eklund, L. *et al.* Mapping quantitative trait loci for carcass and meat quality traits in a wild boar × Large White intercross. *J. Anim. Sci.* **76**, 694–700 (1998).
41. Rubin, C.-J. *et al.* Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* **109**, 19529–19536 (2012).
42. Karim, L. *et al.* Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. *Nat. Genet.* **43**, 405–413 (2011).
43. Setoguchi, K. *et al.* Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (*CW-2*) on chromosome 6 and the Ile-442-Met substitution in *NCAPG* as a positional candidate. *BMC Genet.* **10**, 43 (2009).
44. Quartu, M. *et al.* Neurturin, persephin, and artemin in the human pre- and full-term newborn and adult hippocampus and fascia dentata. *Brain Res.* **1041**, 157–166 (2005).
45. Simanainen, U. *et al.* Evidence for increased tissue androgen sensitivity in neurturin knockout mice. *J. Endocrinol.* **218**, 151–163 (2013).
46. Oschipok, L.W., Teh, J., McPhail, L.T. & Tetzlaff, W. Expression of Semaphorin3C in axotomized rodent facial and rubrospinal neurons. *Neurosci. Lett.* **434**, 113–118 (2008).
47. Hernández-Montiel, H.L., Tamariz, E., Sandoval-Minero, M.T. & Varela-Echavarría, A. Semaphorins 3A, 3C, and 3F in mesencephalic dopaminergic axon pathfinding. *J. Comp. Neurol.* **506**, 387–397 (2008).
48. Gonthier, B. *et al.* Functional interaction between matrix metalloproteinase-3 and semaphorin-3C during cortical axonal growth and guidance. *Cereb. Cortex* **17**, 1712–1721 (2007).
49. Ruediger, T. *et al.* Integration of opposing semaphorin guidance cues in cortical axons. *Cereb. Cortex* **23**, 604–614 (2013).
50. Niquille, M. *et al.* Transient neuronal populations are required to guide callosal axons: a role for semaphorin 3C. *PLoS Biol.* **7**, e1000230 (2009).
51. Pasterkamp, R.J., Kolk, S.M., Hellemons, A.J.C.G.M. & Kolodkin, A.L. Expression patterns of semaphorin7A and plexinC1 during rat neural development suggest roles in axon guidance and neuronal migration. *BMC Dev. Biol.* **7**, 98 (2007).
52. Brown, C.B. *et al.* PlexinA2 and semaphorin signaling during cardiac neural crest development. *Development* **128**, 3071–3080 (2001).
53. Ultanir, S.K. *et al.* Chemical genetic identification of NDR1/2 kinase substrates AAK1 and Rabin8 uncovers their roles in dendrite arborization and spine development. *Neuron* **73**, 1127–1142 (2012).
54. Chevallier, J. *et al.* Rab35 regulates neurite outgrowth and cell shape. *FEBS Lett.* **583**, 1096–1101 (2009).
55. Ong, S.H. *et al.* FRS2 proteins recruit intracellular signaling pathways by binding to diverse targets on fibroblast growth factor and nerve growth factor receptors. *Mol. Cell. Biol.* **20**, 979–989 (2000).
56. Sokol, D.K. *et al.* High levels of Alzheimer β-amyloid precursor protein (APP) in children with severely autistic behavior and aggression. *J. Child Neurol.* **21**, 444–449 (2006).
57. Grayton, H.M., Missler, M., Collier, D.A. & Fernandes, C. Altered social behaviours in neurexin 1α knockout mice resemble core symptoms in neurodevelopmental disorders. *PLoS ONE* **8**, e67114 (2013).

58. Bhoiwala, D.L. *et al.* Overexpression of RCAN1 isoform 4 in mouse neurons leads to a moderate behavioral impairment. *Neurol. Res.* **35**, 79–89 (2013).
59. Dierssen, M. *et al.* Behavioral characterization of a mouse model overexpressing DSCR1/RCAN1. *PLoS ONE* **6**, e17010 (2011).
60. Kim, K.S., Larsen, N., Short, T., Plastow, G. & Rothschild, M.F. A missense variant of the porcine melanocortin-4 receptor (*MC4R*) gene is associated with fatness, growth, and feed intake traits. *Mamm. Genome* **11**, 131–135 (2000).
61. Xu, P. *et al.* Double deletion of melanocortin 4 receptors and SAPAP3 corrects compulsive behavior and obesity in mice. *Proc. Natl. Acad. Sci. USA* **110**, 10759–10764 (2013).
62. Valette, M. *et al.* Eating behaviour in obese patients with melanocortin-4 receptor mutations: a literature review. *Int. J. Obes. (Lond.)* **37**, 1027–1035 (2013).
63. Wojcik, S.M. *et al.* Genetic markers of a Munc13 protein family member, BAIAP3, are gender specifically associated with anxiety and benzodiazepine abuse in mice and humans. *Mol. Med.* **19**, 135–148 (2013).
64. Garud, N.R., Messer, P.W., Buzbas, E.O. & Petrov, D.A. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* **11**, e1005004 (2015).
65. Freedman, A.H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10**, e1004016 (2014).
66. Warmuth, V. *et al.* Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci. USA* **109**, 8202–8206 (2012).
67. Turner, T.L., Hahn, M.W. & Nuzhdin, S.V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, e285 (2005).
68. Ludwig, A. *et al.* Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Phil. Trans. R. Soc. Lond. B* **370**, 20130386 (2015).
69. Vilà, C., Seddon, J. & Ellegren, H. Genes of domestic mammals augmented by backcrossing with wild ancestors. *Trends Genet.* **21**, 214–218 (2005).
70. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
71. Pickrell, J.K. & Reich, D. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* **30**, 377–389 (2014).
72. Larson, G. & Burger, J. A population genetics view of animal domestication. *Trends Genet.* **29**, 197–205 (2013).
73. Girdland Flink, L. *et al.* Establishing the validity of domestication genes using DNA from ancient chickens. *Proc. Natl. Acad. Sci. USA* **111**, 6184–6189 (2014).
74. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

ONLINE METHODS

Sample collection and DNA preparation. Blood samples were collected from a total of 622 individuals comprising 403 European domestic pigs, 92 Asian domestic pigs, 103 European wild boars and 23 Asian wild boars (including a Tibetan wild boar[36]) and a Javan Warty pig (*S. verrucosus*), used as an outgroup[33]. For a full description of the samples, see **Supplementary Table 1**. DNA was extracted from the blood samples with QIAamp DNA Blood Spin kits (Qiagen). The quality and quantity of the DNA extracted were checked on a Qubit 2.0 fluorometer (Invitrogen). SNP genotyping was performed with the Illumina Porcine60K iSelect BeadChip. This data set (60K) was solely used for the PCA and TreeMix analyses displayed in **Supplementary Figures 3, 6** and **7** (described in the **Supplementary Note**). For genome resequencing, we used 1–3 μg of genomic DNA to construct libraries (insert size range of 300–500 bp). Library preparation was conducted according to the Illumina library preparation protocol. Sequencing was carried out on an Illumina HiSeq platform with the 100- and 150-bp paired-end sequencing kits. The resequencing data were used for modeling (ABC; **Supplementary Note**), the selection scan analyses and the PCA, phylogeny and ADMXITURE analyses displayed in **Figures 2** and **3**. DNA was obtained from blood samples collected by veterinarians according to national legislation or from tissue samples from animals obtained from slaughterhouses or, in the case of wild boar, from animals culled within wildlife management programs.

Alignment and variant calling. All samples selected for genome sequencing were sequenced to approximately 10× coverage (**Supplementary Table 1**). Reads were trimmed for minimum Phred quality >20 over three consecutive basepairs and discarded if shorter than 45 bp. Alignment was performed with the Mosaik Aligner (v.1.1.0017) with the unique alignment option to porcine reference genome build 10.2. Variants were called using Genome Analysis Toolkit (GATK) UnifiedGenotyper version 2.8 (ref. 75). We used a prior of 0.01 for the probability of heterozygous calls[34,76].

Approximate Bayesian computation. We used 104 genomes for the ABC analysis. Simulations were performed on 100 10-kb unlinked loci. Backward coalescent simulations with recombination were performed using ms[77] under seven models (**Supplementary Fig. 2**). For model-testing purposes, we ran 200,000 simulations per model. Summary statistics were computed on observed and simulated data using libsequence[78]. We compared all models simultaneously[79] using a standard ABC-GLM general linear model approach as implemented in ABCtoolbox[80]. For parameter inference, we ran 2,000,000 simulations under the best-fitting model. We extracted 10 partial least-square (PLS) components (**Supplementary Fig. 11**) from the 93 summary statistics in the observed and simulated data[81]. We retained a total of 10,000 simulations that were closest to the observed data and applied a standard ABC-GLM approach[82].

Ancestry of European and Asian pigs. We used TreeMix[83] to build a maximum-likelihood population tree from the 60K SNP data set. We generated ten replicates (with different seeds) and selected the run with the highest likelihood score. PCA was performed using flashpca[84] for both the genome-wide (using 500,000 randomly selected SNPs; **Supplementary Note**) and 60K SNP data sets. We built a neighbor-joining tree based on an IBS distance matrix computed with PLINK[84] and the R package ape[85] (**Fig. 2b**), using the same 500,000 randomly selected SNPs. We also used ADMIXTURE[74] to estimate the optimum number of clusters in our data set and to assess the ancestry of our 103 samples also using the same 500,000 SNPs (**Supplementary Note**).

Selection scan. We used SweeD to detect sweeps[38]. To obtain critical threshold values (*P* values), we used a posterior-predictive simulation (PPS) approach. We simulated 2 replicates of 3 Mb each using the parameters of the 10,000 closest retained simulations from our ABC analysis (20,000 simulations). Simulations were run using MACS[86]. We derived a critical threshold for observed CLR values in each population using the CDF derived from the CLR distribution that was obtained from the PPS approach. All regions with $P < 0.01$ were selected for further analysis.

75. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
76. Frantz, L.A.F. *et al.* Evolution of Tibetan wild boars. *Nat. Genet.* **47**, 188–189 (2015).
77. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338 (2002).
78. Thornton, K. libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**, 2325–2327 (2003).
79. Peter, B.M., Huerta-Sanchez, E. & Nielsen, R. Distinguishing between selective sweeps from standing variation and from a *de novo* mutation. *PLoS Genet.* **8**, e1003011 (2012).
80. Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**, 116 (2010).
81. Wegmann, D., Leuenberger, C. & Excoffier, L. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**, 1207–1218 (2009).
82. Leuenberger, C. & Wegmann, D. Bayesian computation and model selection without likelihoods. *Genetics* **184**, 243–252 (2010).
83. Pickrell, J.K. & Pritchard, J.K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
84. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* **9**, e93766 (2014).
85. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
86. Chen, G.K., Marjoram, P. & Wall, J.D. Fast and flexible simulation of DNA sequence data. *Genome Res.* **19**, 136–142 (2009).