

1 **BIOLOGICAL SCIENCES: Evolution**

2  
3 **Analyses of Eurasian wild and domestic pig genomes reveal long-term gene-flow and selection**  
4 **during domestication**

5  
6 Laurent A.F. Frantz<sup>1</sup>, Joshua G. Schraiber<sup>2,3</sup>, Ole Madsen<sup>1</sup>, Hendrik-Jan Megens<sup>1</sup>, Alex Cagan<sup>4</sup>, Mirte  
7 Bosse<sup>1</sup>, Yogesh Paudel<sup>1</sup>, Richard PMA Crooijmans<sup>1</sup>, Greger Larson<sup>5,6</sup> and Martien AM Groenen<sup>1</sup>

8  
9 1 Animal Breeding and Genomics Group, Wageningen University, De Elst 1, Wageningen, WD 6708,  
10 The Netherlands

11 2 Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA.

12 3 Department of Genome Sciences, University of Washington, Seattle, WA 98195-5065, USA

13 4 Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103  
14 Leipzig, Germany.

15 5 Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham DH1  
16 3LE, UK.

17 6 current address: Research Laboratory for Archaeology, Dyson Perrins Building, South Parks Road,  
18 Oxford OX1 3QY, UK

19  
20  
21 Corresponding author: [laurent.frantz@wur.nl](mailto:laurent.frantz@wur.nl)

22  
23 Keywords: Domestication; Gene-flow; ABC; Genomics;

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

## 1 **Significance Statement**

2 Though animal domestication has traditionally been viewed as a human-directed process involving  
3 small populations of domestic animals and limited interbreeding between wild and domestic forms, but  
4 recent zooarcheological insights have questioned this model. By studying domestication in pigs, we  
5 demonstrate that complex models of domestication incorporating long-term gene-flow from multiple  
6 wild boar populations fit genomic data from modern wild and domestic pigs significantly better than  
7 models based on the traditional perspective. In addition, we demonstrate that selection at genes  
8 associated with domestic traits countered the effects of the gene flow, thus allowing morphological and  
9 behavioural differentiation between wild and domestic populations to be maintained.

10

## 11 **Abstract**

12 Traditionally, the process of domestication is assumed to be initiated by people, involve few individuals  
13 and rely on reproductive isolation between wild and domestic forms. However, an emerging  
14 zooarcheological consensus depicts animal domestication as a long-term process without reproductive  
15 isolation or strong intentional selection. Here, we ask whether pig domestication followed a traditional  
16 linear model, or a complex, reticulate model as predicted by zooarcheologists. To do so, we fit models  
17 of domestication to whole genome data from over 100 wild and domestic pigs. We found that the  
18 assumptions of traditional models, such as reproductive isolation and strong domestication bottlenecks,  
19 are incompatible with the genetic data and provide support for the zooarcheological theory of a  
20 complex domestication process. In particular, gene-flow from wild to domestic pigs was a ubiquitous  
21 feature of the domestication of pigs. In addition, we show that despite gene-flow, the genomes of  
22 domestic pigs show strong signatures of selection at loci that affect behaviour and morphology.  
23 Specifically, our results are consistent with independent parallel sweeps in two independent

1 domestication areas (China and Anatolia) at loci linked to morphological traits. We argue that recurrent  
2 selection for domestic traits likely counteracted the homogenising effect of gene-flow from wild boars  
3 and created “islands of domestication” in the genome. Overall, our results suggest that genomic  
4 approaches that allow for more complex models of domestication to be embraced should be employed.  
5 The results from these studies will have significant ramifications for studies that attempt to infer the  
6 origin of domesticated animals.

7 \body

## 8 **Introduction**

9 The rise of agriculture, which occurred approximately 10,000 years ago, was one of the most important  
10 transitions in human history. During the Neolithic revolution, the domestication of plant and animal  
11 species led to a major subsistence shift, from hunter-gatherers to sedentary agriculturalists that  
12 ultimately resulted in the development of complex societies. The process of animal domestication led  
13 to striking morphological and behavioural changes in domesticated organisms compared with their wild  
14 progenitors (1). Traditionally, this process has often been viewed as human-directed, involving strong  
15 bottlenecks in the domestic population (*i.e.* founder events due to the selection of only a few  
16 individuals at the beginning of domestication) and reproductive isolation between wild and domestic  
17 forms (2–6). This straightforward model provides an attractive theoretical framework for geneticists,  
18 because key events such as the geographic origin and timeframe of domestication are well defined.  
19 Thus, the assumption of reproductive isolation eases the interpretation of genetic data from domestic  
20 and wild forms. For instance, under this model, geneticists have interpreted phylogenetic affinities of  
21 domestic animals with multiple, geographically divergent wild populations as evidence of frequent,  
22 independent domestication origins in multiple species (8–13).

1 However, this view conflicts with zooarcheological evidence that shows that domestication episodes  
2 are rare, and that domesticated forms were diffused out from a limited number of core regions (7, 14,  
3 15). Moreover, there is a growing body of empirical and theoretical archaeological work (3, 4, 16) that  
4 challenges the simplicity of traditional models. In these new, more complex models, pre-historic  
5 domestication of animals is viewed as mainly unintentional (3, 4, 7) and neither reproductive isolation  
6 nor strong intentional selection are viewed to be as crucial and widespread as previously thought.  
7 Instead, domestication is seen as a long-term, diffuse process (17), involving gene-flow (during as well  
8 as post-domestication) between wild and domestic forms (18) and with emphases on multiple, taxon  
9 specific, human-animal relationships (3, 4). The possibility of post-domestication gene-flow between  
10 domestic animals and their wild progenitors, as well as a lack of strong domestication bottlenecks (18),  
11 are key predictions from this novel framework that contrast with more traditional models of  
12 domestication. Moreover, extensive gene-flow between wild and domestic forms violates the  
13 assumptions of traditional models of domestication and has significant ramifications for studies that  
14 attempt to infer the spatial and chronological origin of domestication using genetic data.

15 Here, we focus on pig domestication using genome-wide datasets of modern domestic pigs and wild  
16 boars. Pigs were domesticated independently once in Anatolia (16) and once in the Mekong valley  
17 around 9,000 BP (19). Furthermore, ancient mtDNA analyses found that the first domestic pigs in  
18 Europe were transported by early farmers from the Levant into Europe around 5,500 BC, concordant  
19 with zooarchaeological evidence for a single domestication origin of Western Eurasian domestic pigs  
20 (20, 21). However, a few thousand years after their introduction, domestic pigs in Europe had  
21 completely lost the Near Eastern mtDNA signatures and instead acquired mtDNA haplotypes typically  
22 found in local European wild boars (20, 21). These findings suggest that early domestic populations

1 experienced post-domestication gene-flow from wild boar populations that were not involved in the  
2 Anatolian domestication process (7).

3 Further mtDNA analyses of ancient Anatolian material demonstrated that, by 500 BC, local mtDNA  
4 haplotypes were also replaced by haplotypes from European wild boars. This result suggests extensive  
5 mobile swineherding throughout Europe and Anatolia (21), consistent with both archaeological and  
6 historical evidence, as well as limited management and selection up until the industrial revolution in the  
7 19th century (22, 23). Thus, under a complex model of domestication, mtDNA replacement in ancient  
8 European and Anatolian pigs is the result of post-domestication gene-flow, loose pig management and  
9 mobile swine herding. We therefore expect such phenomenon to have left a strong signal of gene-flow  
10 from wild boars in the genome of modern domestic pigs.

11 However, while unsupported by any zooarcheological evidence, the observed mtDNA turnovers could  
12 also be interpreted as a *de-novo* domestication of a population of European wild boars rather than the  
13 result of post-domestication gene-flow from wild boars. Moreover, because of its mode of inheritance  
14 and limited resolution, small mtDNA markers provide a very limited impression of gene-flow, making  
15 it impossible to test these hypotheses. Thus, the hypothesis of complex domestication in pigs has yet to  
16 be tested with the resolution and confidence afforded by unlinked, nuclear markers. In addition, unlike  
17 horses and donkeys, intentional interbreeding between pigs and wild boars confers no clear productive  
18 advantage and is thought as being mainly unintentional (18). Lastly, there is a clear morphological and  
19 behavioural dichotomy between wild and domestic pigs that is evident in modern animals as well as in  
20 the zooarcheologic record (24–27). Thus, the possibility of unintentional gene-flow between wild and  
21 domestic pigs also raises questions regarding the mechanisms behind the maintenance of traits that  
22 differentiate domestic and wild forms.

1 Here, we fit models of domestication to a genome-wide dataset from over 100 wild and domestic pigs.  
2 Our main aim is to ask whether pig domestication follows a traditional, linear model or a complex,  
3 reticulate model. More precisely, we assessed whether the zooarcheological evidence for a single,  
4 geographically restricted, domestication of (Western) pigs in Anatolia (7, 15, 20) is compatible with the  
5 assumption of a traditional model of domestication involving reproductive isolation and strong  
6 bottlenecks.

## 7 **Results and Discussion**

8 We evaluated the support of multiple models for the domestication of pigs in Europe and Asia. Our  
9 analysis focused on 103 genomes from European wild boars (EUW) (8), European commercial /  
10 historical domestic (EUD) (Table S1). In addition, this data set comprises multiple populations of Asian  
11 wild boar (ASW) and Asian domestic pigs (ASD; Table S1). In order to better understand the early  
12 process of domestication, we sampled a range of wild boar populations, from Asia and all major  
13 European Pleistocene refugia, rare historical European and Asian breeds, as well as modern  
14 commercial breeds. To test key predictions of the complex domestication framework described above,  
15 we fit simple but informative models to these genomic data set using Approximate Bayesian  
16 Computation (ABC) (see Materials and Methods).

### 17 *Testing models of domestication from genome sequences*

18 We first tested the hypothesis of gene-flow between wild and domestic pigs. More precisely, we asked  
19 whether reproductive isolation between wild and domestic pigs is compatible with the zooarcheologic  
20 evidence that pigs were domesticated only twice, independently in Anatolia and China. To do so, we  
21 first evaluated the fit of the traditional model in which domestication is modelled as two parallel events  
22 in Asia and Europe. In this model, domestication takes place at time T1 in Europe and T2 in Asia and  
23 involves no gene-flow between wild and domestics (reproductive isolation) or between domestics from

1 Asia and Europe (Fig. 1a). We then compared this null model to five other models involving different  
2 patterns of continuous mixture: within wild, within domestic, between wild and domestic, etc. (Fig.  
3 S1). By comparing these six models (Fig. S1), we found that a model involving gene-flow between  
4 domestic and wild (within Asia and Europe) as well as between domestic and domestic (between  
5 Europe and Asia) provided a large improvement of fit (Bayes Factor [BF] > 14) when compared to any  
6 other model tested in this study (Fig. 1a; Fig. S1). Thus, our explicit modelling framework provides  
7 very strong evidence that reproductive isolation between wild and domestics was not maintained during  
8 and after domestication in Asia and Europe.

9 We further assessed this finding using a dataset of over 600 pigs (from the same populations as in the  
10 genome-wide data) that were genotyped on the Porcine SNP60 array (SI Text). We investigated the  
11 historical relationship among these populations using *TreeMix* (28). Our analysis showed that EUD and  
12 ASD were both paraphyletic while EUW was monophyletic (SI Text; Fig. S2). The paraphyly of EUD  
13 and ASD is difficult to reconcile with the assumptions of a linear, spatially restricted model of  
14 domestication. Instead, this finding provides further evidence of a complex domestication process that  
15 involved gene-flow between wild and domestic pigs. Moreover, we found that gene-flow between wild  
16 and domestic pigs in Europe was strongly asymmetrical (migration rate from EUW to EUD was much  
17 higher than from EUD to EUW; SI Text; Fig. S3)

18 Lastly, we also found evidence that Asian and European domestic pigs exchanged genetic material.  
19 This is consistent with previous studies and is most likely the result of European importation of  
20 Chinese pigs during the 19th century to improve European commercial breeds (23, 29, 30). However,  
21 gene-flow between domestic populations (EUD and ASD) was very limited relative to gene-flow  
22 between wild and domestics (EUW and EUD; ASW and ASD; Fig. S3). This is not surprising given our  
23 sampling strategy that focuses on historical (non-commercial) European domestic breeds that are less

1 likely to be admixed with Asian domestic pigs (23). The small amount of gene-flow between domestic  
2 pigs suggests that admixture between domestic pigs had no little influence on our conclusion of gene-  
3 flow between wild and domestics (see SI Text).

4 Together, these findings demonstrate that domestic pigs do not form a homogeneous genetic group, as  
5 expected under a simple human-driven model of domestication. Instead, domestic pigs are a genetic  
6 mosaic of different wild boar populations. Thus, modern genetic data from domestic pigs can only be  
7 reconciled with zooarcheological evidence for a restricted domestication process if modelled with  
8 continuous gene-flow between wild and domestic pigs.

### 9 *Demography of pig domestication*

10 We also tested for the presence of a strong population bottleneck associated with domestication. To do  
11 so, we estimated the posterior distribution of demographic parameters using 10,000 retained  
12 simulations out of 2,000,000. Under the assumption of a simple linear model of domestication with no  
13 gene-flow and strong intentional selection by humans, we would expect a strong bottleneck in domestic  
14 populations. Overall, we found a population decline in both EUW and EUD (Fig. 2). This is consistent  
15 with previous results demonstrating that Pleistocene glaciations resulted in long-term population  
16 decline in European wild boars (23, 31–33). However, this population decline was more pronounced in  
17 EUW than in EUD (Fig 2). In addition, we found that the effective population size of EUD ( $N_e$ -EUD  
18  $\approx 20,563$ ) was more than twice as large as the effective population size of EUW ( $N_e$ -EUW  $\approx 8,497$ ).  
19 This is most likely due to a series of strong bottlenecks in the European wild population, caused by  
20 over-hunting and loss of suitable habitat (23, 31-33). Together these results do not support the existence  
21 of a strong domestication bottleneck in European domestic pigs and instead support the contention that  
22 continuous gene-flow from multiple genetically and geographically distinct wild boar populations  
23 likely increased the effective population size of EUD.



1 *Gene-flow from wild boar populations not involved in the domestication process*

2 We showed that a model incorporating continuous gene-flow between wild and domestic pigs is  
3 significantly more compatible zooarcheological evidence compared to a traditional hypothesis of  
4 reproductive isolation. Despite this fact, we only modelled gene-flow from a population of wild boars  
5 that we assumed to be close to the source of domestication. Here, we test the hypothesis that another  
6 population of wild boars either extinct (due to hunting pressure and habitat loss (23)) or un-sampled in  
7 our analysis (*e.g.* from Central Eurasia, where we did not sample) also contributed to the gene-pool of  
8 modern European domestic pigs. To do so, we used a model that is similar to our best fitting model (see  
9 above; Fig. 1a) but with an additional “ghost” population that splits from EUW/EUD during the  
10 Pleistocene (Fig. 1b) and act as a step between ASW and EUW (migration ASW  $\leftrightarrow$  Ghost  $\leftrightarrow$  EUW  
11 but also Ghost  $\leftrightarrow$  EUD; Fig. 1b). This model provided a substantial improvement of fit (BF>6). This  
12 result shows that mobile herding of domestic pigs across Europe most likely resulted in gene-flow from  
13 a least one wild boar population that was genetically divergent from the population involved in the  
14 domestication process in Anatolia.

15 *Positive selection in domestic pigs*

16 Our analysis shows that gene-flow between wild and domestic forms was a ubiquitous feature of  
17 domestication and post-domestication processes in pigs. Thus, extensive gene-flow from wild boars  
18 into domestic pigs during and after domestication raises questions regarding the mechanisms behind  
19 the maintenance of the clear morphological and behavioural differences observed between domestic  
20 and wild pigs. Intentional or unintentional artificial selection could have counteracted the effect of  
21 gene-flow and resulted in morphological and behavioural differentiation between wild and domestics.  
22 In order to assess the importance of selection in the genome of domestic pigs in face of gene-flow we  
23 conducted a scan for positive selection using *SweeD* (34, 35). *SweeD* computes the composite

1 likelihood ratio (CLR) of a sweep model over a neutral model. Such a test can be very sensitive to  
2 demography and migration (36). To correct for effects of demography and migrations we used the  
3 10,000 closest simulations (out of 2,000,000) under our best fitting model in Fig. 1a (without ghost  
4 population, see SI Text) to generate an expected cumulative distribution function (CDF) of neutral CLR  
5 and to compute the p-value for all empirical CLR value in the genome (Materials and Methods; SI  
6 Text). We identified 249 and 136 10kbp regions with  $p < 0.01$  in the genome of European and Asian  
7 domestics, respectively.

8 First, we examined sweeps private to each population (SI Text). These sweeps in domestic pigs (EUD  
9 and ASD) were significantly enriched with GO terms related to multiple developmental processes of  
10 bones, teeth and nervous system (Table S1&S2). These terms comprise multiple gene candidates  
11 related to height (*PLAG1*, *NCPAG*, *PENK*, *RPS20* and *LYN* in EUD; Fig. 3a; *LEMD3* and *UPK1* in  
12 ASD) in pigs (37, 38) and cattle (39, 40), nervous system development and maintenance (*NRTN*,  
13 *SEMA3C*, *PLXNC1*, *AAKI*, *RAB35*, *FRS2*) (41–52) as well as genes directly influencing behaviour (*i.e.*  
14 aggressiveness and feeding; *APBA2*, *MC4R*, *RCANI*, *BAIPA3*) (53–60). These results suggest that  
15 domestication and/or post-domestication selection for behavioural and morphological traits was  
16 important in Asian and European domestic pigs and most likely counteracted the effect of continuous  
17 gene-flow in certain parts of the genome.

18 However, the mechanism behind this maintenance remains unknown. One possibility is that there was  
19 recurrent selection for similar traits. This phenomenon may have resulted in parallel sweeps at the same  
20 loci in both ASD and EUD. To investigate this possibility, we looked for signals of parallel selection  
21 between the two independent domestication episodes (ASD and EUD). To do so, we identified genes  
22 with CLR above the significance threshold in both ASD and EUD but not in ASW and EUW. In order  
23 to rule out admixture between ASD and EUD as the cause for observing overlapping significant signal

1 we conducted a phylogenetic analysis in each region separately (SI Text). The genealogy of some of  
2 these regions shows a signal that is consistent with introgression between EUD and ASD (*e.g.* Fig. S7).  
3 However, we found one region of particular interest seems to have swept independently in EUD and  
4 ASD (Fig. 3). Phylogenetic analysis in this region (Fig. 3b) reveals that ASD and ASW as well as EUD  
5 and EUW are mostly monophyletic (Fig. 3c), suggesting an independent sweep in ASD and EUD.  
6 Interestingly, while this sweep does not overlap with genes, the region is just a few kbp upstream of the  
7 highest CLR in EUD (among others; Fig. 3a). This region has been shown to have a strong effect on  
8 body height and stature in pigs (37, 38) and cattle (39, 40). In particular, variation in this region  
9 explains up to 18% in body length difference between wild boars and commercial EUD (37). Given the  
10 importance of this region for morphology in commercial EUD (38) it is possible that human-mediated  
11 selection for similar traits in Asian and European domestic pigs resulted in parallel sweeps at the same  
12 loci. Parallel selection of this form may be the responsible for some of the morphological convergence  
13 in the two independent domestication events in Europe and Asia. Thus, while the phenotypic effect of  
14 this sweep is still unclear, this region provides a particularly interesting candidate to further study the  
15 possibility of convergence between ASD and EUD.

## 16 **Conclusions**

17 The generation of larger amounts of genomic data with ever-greater resolution is allowing us to  
18 embrace the complexity of domestication. The commensurate advancements in theoretical and  
19 empirical perspectives is allowing for more sophisticated models to be tested and for a greater  
20 understanding of animal domestication. In this study we demonstrated that the assumptions of  
21 traditional models, such as reproductive isolation and strong domestication bottlenecks, are  
22 incompatible with the zooarcheological evidence of a geographically restricted domestication process  
23 in pigs. Instead our model testing approach revealed that continuous gene-flow from wild boars to

1 domestic pigs is necessary to reconcile modern genetic data with the zooarcheological evidence.  
2 Moreover, we demonstrated that in Western Eurasia, gene-flow most likely involved at least a second,  
3 un-sampled (possibly even extinct) population of wild boars that was quite divergent from the source of  
4 domestication. This is most likely the result of mobile domestic swine herding, as predicted by  
5 zooarcheologists (18, 21).

6 Such extensive gene-flow from wild boars raises questions regarding the maintenance of morphological  
7 and behavioural traits in domestic pigs. Our study reveals extensive evidence of selection at candidate  
8 genes that influence anatomical and nervous system development, suggesting that selection may have  
9 counteracted the homogenizing effect of gene-flow and maintained the genetic basis for the  
10 morphological and behavioural dichotomy observed between wild and domestic pigs. In addition, our  
11 results show that regions close to genes governing morphological traits have been subject to selection  
12 in parallel in Asia and Europe. In the context of speciation, studies focusing on systems in which gene-  
13 flow is common have identified genomic regions that show excessive inter-specific divergence (*e.g.*,  
14 61). These studies have suggested that such regions may be resistant to gene-flow and likely allowed  
15 for the maintenance of species characteristics (genomic “islands of speciation”) (61). Here we  
16 hypothesise that an analogous process is taking place during pig domestication. By recurrently  
17 selection for similar traits through artificial selection, but allowing for gene-flow, farmers have created  
18 genomic “island of domestication”, that we define as genomic regions governing domestic traits and  
19 thus less affected by gene-flow from wild boars. However, it is unclear whether these sweeps involved  
20 recurrent selection of different haplotypes from standing genetic variation in wild boars or are the result  
21 of selection from *de-novo* mutations. Thus, our results highlight a list of candidate genes that will  
22 provide further studies with the means to further test these hypotheses.

1 Lastly, it is important to underline the limitations of modern DNA and traditional domestication models  
2 to determine the origin and time of domestication of animals, as well as to identify the genes involved  
3 during domestication. Indeed, extensive gene-flow clearly violates the assumption of traditional models  
4 and likely eroded most of the signal to infer time and geographic parameters (62-64). It is therefore  
5 important to apply caution when conducting comparative analyses of modern genetic material from  
6 wild and domestic animals. However, future sequencing of ancient DNA, together with more realistic  
7 modelling frameworks, such as the one presented here, will provide the necessary information not only  
8 to determine the origin and time of domestication of animals but also to identify genes involved during  
9 domestication and will ultimately significantly enhance our knowledge of this fascinating evolutionary  
10 process.

## 11 **Materials and Methods**

### 12 *Sample collection and DNA preparation.*

13 Blood samples were collected from a total of 622 individuals, 403 European domestics, 92 Asian  
14 domestics, 103 European wild boars and 23 Asian wild boars and a Javan Warty pig (*S. verrucosus*),  
15 used as an outgroup (33). For full description of the samples see Table S1. DNA was extracted from the  
16 blood samples with QIAamp DNA blood spin kits (Qiagen Sciences). Quality and quantity of DNA  
17 extraction was checked on a Qubit 2.0 fluorometer (Invitrogen). Single nucleotide polymorphism  
18 (SNP) genotyping was performed with the Illumina Porcine 60K iSelect Beadchip. For the genome re-  
19 sequencing, we used 1–3 ug of genomic DNA to construct libraries (insert size range 300–500 bp).  
20 Library preparation was conducted according to the Illumina library preparation protocol (Illumina  
21 Inc.). Sequencing was done on Illumina Hi-Seq with 100 and 150 paired-end sequencing kits.

### 22 *Alignment and variant calling*

1 All samples selected for genome sequencing were sequenced to approximately 10x coverage (Table  
2 S1). Reads were trimmed for a minimum phred quality  $> 20$  over three consecutive base pairs and  
3 discarded if shorter than 45 base pairs. Alignment was performed with Mosaik Aligner (V. 1.1.0017)  
4 with the unique alignment option to the Porcine reference genome build 10.2. Variants were called  
5 using GATK Unified Genotyper version 2.8 (65). We used a prior of 0.01 for the probability of  
6 heterozygous calls (32).

### 7 *Approximate Bayesian Computation*

8 104 genomes were used for the ABC analysis. Simulations were performed on 100 10kbp unlinked  
9 loci. Backward coalescent simulations with recombination were performed using *ms* (66) under 7  
10 models (Fig. S1). For model testing purposes, we ran 200,000 simulations per model. Summary  
11 statistics were computed on observed and simulated data using *libsequence* (67). We compared  
12 all models simultaneously (68) using a standard ABC-GLM approach as implemented in *ABCtoolbox*  
13 (69). For parameters inference we ran 2,000,000 simulations under the best fitting model. We extracted  
14 10 Partial Least Square (PLS) components from the 93 summary statistics in the observed and  
15 simulated data (70). We retained a total of 10,000 simulations closest to the observed data and applied a  
16 standard ABC-GLM (71).

### 17 *Exploratory analysis using SNP array*

18 We used *TreeMix* (28) to build a maximum likelihood (ML) population tree from the 60K SNP dataset.  
19 We generated 10 replicates (with different seeds) and selected the run with the highest likelihood score.  
20 The PCA analysis was performed using *flashpca* (72).

### 21 *Selection scan*

22 We used SweeD to detect sweeps (35). To obtain critical threshold values (p-values), we used a  
23 posterior predictive simulation (PPS) approach. We simulated 2 replicates of 3Mbp each using the

1 parameters of the 10,000 closest retained simulations from our ABC analysis (20,000 simulations).  
2 Simulations were run using macs (73). We derived a critical threshold for observed CLR in each  
3 population using the cumulative descriptive function (CDF) derived from the CLR distribution that was  
4 obtained from the PPS. All regions with  $p < 0.01$  were selected for further analysis.

## 5 6 **Acknowledgments**

7 We thank Ben Peter for his help and guidance during the model-fitting step of the analysis as well as  
8 for kindly sharing his code. We are also indebted to Daniel Wegmann for providing us with the latest  
9 version of *ABCtoolbox*. We also thank Konrad Lohse for his insights during the conception of the  
10 project. This project is financially supported by the European Research Council under the European  
11 Community's 256 Seventh Framework Programme (FP7/2007–2013)/ERC Grant agreement no  
12 249894. JGS was supported by National Institutes of Health grant R01-40282 and National Science  
13 Foundation postdoctoral fellowship DBI-1402120.

## 14 15 **References**

- 16 1. Darwin C (1868) *The Variation of Animals and Plants Under Domestication* (Murray, John, London).
2. Price EO (2002) *Animal Domestication and Behavior* (CABI Publishing, New York).
3. Zeder MA (2011) in Harlan II: *Biodiversity in Agriculture: Domestication, Evolution and Sustainability*, eds Damania A, Gepts P (Univ California Press, Davis), pp 227-229.
4. Vigne J-D (2011) The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies* 334:171-81.
5. Driscoll CA, Macdonald DW, O'Brien SJ (2009) From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl :9971-8.
6. O'Connor TP (2007) Wild or domestic? Biometric variation in the cat *Felis silvestris* Schreber. *International Journal of Osteoarchaeology* 17:581-595.
7. Larson G, Fuller DQ (2014) *The Evolution of Animal Domestication*. *Annual Review of Ecology, Evolution and Systematics* in press.
8. Larson G et al. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science (New York, NY)* 307:1618-21.
9. Hanotte O et al. (2002) African pastoralism: genetic imprints of origins and migrations. *Science (New York, NY)* 296:336-9.
10. Luikart G et al. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America* 98:5927-32.
11. Naderi S et al. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America* 105:17659-64.

12. Pedrosa S et al. (2005) Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proceedings Biological sciences / The Royal Society* 272:2211-7.
13. Vilà C et al. (2001) Widespread origins of domestic horse lineages. *Science (New York, NY)* 291:474-7.
14. Macneish RS (1992) *The origins of agriculture and settled life* ed University of Oklahoma Press (Norman).
15. Zeder MA (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America* 105:11597-604.
16. Eryvynck A, Hongo H, Dobney K, Meadow R (2001) Born Free ? New Evidence for the Status of *Sus scrofa* at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient* 27:47-73.
17. Dobney K, Larson G (2006) Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* 269:060222013030001-???
18. Marshall FB, Dobney K, Denham T, Capriles JM (2014) Evaluating the roles of directed breeding and gene flow in animal domestication. *Proceedings of the National Academy of Sciences of the United States of America* 111:6153-8.
19. Cucchi T, Hulme-Beaman A, Yuan J, Dobney K (2011) Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. *Journal of Archaeological Science* 38:11-22.
20. Larson G et al. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America* 104:15276-81.
21. Ottoni C et al. (2013) Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. *Molecular biology and evolution*:mss261-.
22. Albarella U, Manconi F, Trentacoste A (2011) in *Ethnozooarchaeology. The present and past of human-animal relationships*, eds Albarella U, Trentacoste A (Oxbow Books, Oxford), pp 143-159.
23. White S (2011) From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History. *Environmental History* 16:94-120.
24. Owen J et al. (2014) The zooarchaeological application of quantifying cranial shape differences in wild boar and domestic pigs (*Sus scrofa*) using 3D geometric morphometrics. *Journal of Archaeological Science* 43:159-167.
25. Evin A et al. (2013) The long and winding road: identifying pig domestication through molar size and shape. *Journal of Archaeological Science* 40:735-743.
26. Albarella U, Davis SJM, Detry CP, Rowley-Conwy P (2005) Pigs of the “ Far West ”: the biometry of *Sus* from archaeological sites in Portugal. *Anthropozoologica* 40:27-54.
27. Rowley-Conwy P, Albarella U, Dobney K (2012) Distinguishing Wild Boar from Domestic Pigs in Prehistory: A Review of Approaches and Recent Results. *Journal of World Prehistory* 25:1-44.
28. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide



allele frequency data. *PLoS genetics* 8:e1002967.

29. Groenen MAM et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491:393-398.

30. Bosse M et al. (2014) Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nature communications* 5:4392.

31. Groenen MAM et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491:393-8.

32. Bosse M et al. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS genetics* 8:e1003100.

33. Frantz LA et al. (2013) Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome biology* 14:R107.

34. Nielsen R et al. (2005) Genomic scans for selective sweeps using SNP data. *Genome research* 15:1566-75.

35. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N (2013) SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular biology and evolution* 30:2224-34.

36. Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping It Local: Evidence for Positive Selection in Swedish *Arabidopsis thaliana*. *Molecular biology and evolution*:msu247-.

37. Andersson-Eklund L et al. (1998) Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x Large White intercross. *Journal of animal science* 76:694-700.

38. Rubin C-J et al. (2012) Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences of the United States of America* 109:19529-36.

39. Karim L et al. (2011) Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. *Nature genetics* 43:405-13.

40. Setoguchi K et al. (2009) 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 genetics* 10:43.

41. Quartu M et al. (2005) Neurturin, persephin, and artemin in the human pre- and full-term newborn and adult hippocampus and fascia dentata. *Brain research* 1041:157-66.

42. Simanainen U et al. (2013) Evidence for increased tissue androgen sensitivity in neurturin knockout mice. *The Journal of endocrinology* 218:151-63.

43. Oschipok LW, Teh J, McPhail LT, Tetzlaff W (2008) Expression of Semaphorin3C in axotomized rodent facial and rubrospinal neurons. *Neuroscience letters* 434:113-8.

44. Hernández-Montiel HL, Tamariz E, Sandoval-Minero MT, Varela-Echavarría A (2008) Semaphorins 3A, 3C, and 3F in mesencephalic dopaminergic axon pathfinding. *The Journal of comparative neurology* 506:387-97.

45. Gonthier B et al. (2007) Functional interaction between matrix metalloproteinase-3 and semaphorin-3C during cortical axonal growth and guidance. *Cerebral cortex (New York, NY : 1991)* 17:1712-21.

46. Ruediger T et al. (2013) Integration of opposing semaphorin guidance cues in cortical axons. *Cerebral cortex* (New York, NY : 1991) 23:604-14.
47. Niquille M et al. (2009) Transient neuronal populations are required to guide callosal axons: a role for semaphorin 3C. *PLoS biology* 7:e1000230.
48. Pasterkamp RJ, Kolk SM, Hellemons AJCGM, Kolodkin AL (2007) Expression patterns of semaphorin7A and plexinC1 during rat neural development suggest roles in axon guidance and neuronal migration. *BMC developmental biology* 7:98.
49. Brown CB et al. (2001) PlexinA2 and semaphorin signaling during cardiac neural crest development. *Development* 128:3071-3080.
50. Ultanir SK et al. (2012) Chemical genetic identification of NDR1/2 kinase substrates AAK1 and Rabin8 Uncovers their roles in dendrite arborization and spine development. *Neuron* 73:1127-42.
51. Chevallier J et al. (2009) Rab35 regulates neurite outgrowth and cell shape. *FEBS letters* 583:1096-101.
52. Ong SH et al. (2000) FRS2 Proteins Recruit Intracellular Signaling Pathways by Binding to Diverse Targets on Fibroblast Growth Factor and Nerve Growth Factor Receptors. *Molecular and Cellular Biology* 20:979-989.
53. Sokol DK et al. (2006) High Levels of Alzheimer Beta-Amyloid Precursor Protein (APP) in Children With Severely Autistic Behavior and Aggression. *J Child Neurol* 21:444-449.
54. Grayton HM, Missler M, Collier DA, Fernandes C (2013) Altered social behaviours in neurexin 1 $\alpha$  knockout mice resemble core symptoms in neurodevelopmental disorders. *PloS one* 8:e67114.
55. Bhoiwala DL et al. (2013) Overexpression of RCAN1 isoform 4 in mouse neurons leads to a moderate behavioral impairment. *Neurological research* 35:79-89.
56. Dierssen M et al. (2011) Behavioral characterization of a mouse model overexpressing DSCR1/RCAN1. *PloS one* 6:e17010.
57. Kim KS, Larsen N, Short T, Plastow G, Rothschild MF (2000) A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits. *Mammalian Genome* 11:131-135.
58. Xu P et al. (2013) Double deletion of melanocortin 4 receptors and SAPAP3 corrects compulsive behavior and obesity in mice. *Proceedings of the National Academy of Sciences of the United States of America* 110:10759-64.
59. Valette M et al. (2013) Eating behaviour in obese patients with melanocortin-4 receptor mutations: a literature review. *International journal of obesity* (2005) 37:1027-35.
60. Wojcik SM et al. (2013) Genetic markers of a Munc13 protein family member, BAIAP3, are gender specifically associated with anxiety and benzodiazepine abuse in mice and humans. *Molecular medicine* (Cambridge, Mass) 19:135-48.
61. Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS biology* 3:e285.
62. Pickrell J, Reich D (2014) Towards a new history and geography of human genes informed by ancient DNA (Cold Spring Harbor Labs Journals)

63. Larson G, Burger J (2013) A population genetics view of animal domestication. *Trends in genetics: TIG* 29:197-205.
64. Girdland Flink L et al. (2014) Establishing the validity of domestication genes using DNA from ancient chickens. *Proceedings of the National Academy of Sciences of the United States of America* 111:6184-9.
65. McKenna A et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297-1303.
66. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337-338.
67. Thornton K (2003) libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325-2327.
68. Peter BM, Huerta-Sanchez E, Nielsen R (2012) Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS genetics* 8:e1003011.
69. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC bioinformatics* 11:116.
70. Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182:1207-18.
71. Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* 184:243-52.
72. Abraham G, Inouye M (2014) Fast principal component analysis of large-scale genome-wide data. *PloS one* 9:e93766.
73. Chen GK, Marjoram P, Wall JD (2009) Fast and flexible simulation of DNA sequence data. *Genome research* 19:136-42.

### Figure Legends:

**Figure 1: Schematic representation of models tested in this study.** Double headed arrows show migrations that were modelled as two independent continuous parameters. **a)** Model testing approach comparing six models. Two models, one without gene-flow (null model; top) and one model with gene-flow between wild (ASW and EUW) and domestic (ASD and EUD) as well as between domestic and wild (full; middle) are displayed. The four additional model tested in this study are displayed in Fig. S1. Bayes Factors in this square were computed without the Ghost model (6 models in total; see b)). **b)** Same as a) but with the Ghost model (bottom). Bayes Factors were computed with all 7 models together.

**Figure 2: Posterior density distribution of demographic parameters.** Population sizes are shown in relative population proportion (the ratio of the current population size over the population size before T0 [Figure 1]).

**Figure 3: Example of a parallel sweep in ASD and EUD.** **a)** Composite likelihood ratio (CLR) values in the PLAG1 region. Dashed blue and red lines represent  $p=0.05$  and  $0.01$  thresholds respectively. Grey shaded area is the position of the parallel sweep (see b). **b)** CLR values in the

parallel sweep region a few thousands base pair upstream of the *PLAG1* region. **c)** Genealogy of phased haplotypes for the region in Fig. 2b. Shaded area highlights the very short branch lengths that are the result of a sweep. The shaded area on the left side (Europe) contains 64 haplotypes from EUD (>72% of total EUD haplotypes) and 2 haplotypes from EUW (<4% of total EUW haplotypes). The shaded area on the right side (Asia) contains 24 haplotypes from ASD (>54% of total ASD haplotypes) and no ASW haplotype.





