Supplementary Note for:

Evolution of Tibetan Wild Boars.

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

[1]: *Animal Breeding and Genomics Group, Wageningen University, Wageningen, The Netherlands.*

[2]: *Department of Integrative Biology, University of California, Berkeley, USA.*

[3]: *Department of Genome Sciences, University of Washington, Seattle, WA 98195-5065, USA*

[4]: *Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham, UK.*

[5]: *Present address: Palaeogenomics & Bio-Archaeology Research Network, Research Laboratory for Archaeology & the History of Art, University of Oxford, Oxford OX1 3QY, UK*

Corresponding authors: * laurent.frantz@arch.ox.ac.uk; * martien.groenen@wur.nl;

1. ***Alignment and genotype call.***

We downloaded a single library of pair-end reads (500bp insert-size; accession: SRR652444) from the Tibetan wild boar that was assembled in Li *et. al.* from the NCBI website. Reads were uniquely aligned (reads that mapped to multiple positions in the genome with equal mapping quality were discarded) to the reference genome of *Sus scrofa* (Ssc10.2) using SMALT[1]. Alignment statistics are presented in Table S1.

We called genotypes from BAM files using GATK[4] for 8 Sus samples (including Tibetan Wild boar) and an outgroup (Warthog; Figure 1a), which were previously aligned to Ssc10.2[2,3]. The prior for heterozygous calls was set at 0.01.

2. ***Phylogenetics analysis.***

We first identified loci suitable for phylogenetic analysis. To do so we divided the reference genome of *S. scrofa* (Ssc10.2) into 1kb blocks. For each sample, we filtered out any loci covered by more than twice the genome-wide average depth of coverage or that had more than 10% missing data. Thereafter, we merged neighboring 1k loci, each of which passed our criteria in every sample, into 10kb loci. We then randomly sampled one hundred 10kb loci. These loci were concatenated together in a multi-partition alignment (totaling 1Mbp). We then computed a maximum likelihood (ML) phylogenetic tree with 100 bootstrap replicates using RAxML[5], fitting a separate GTR+G4 model of substitution to each partition. We repeated the concatenation analysis, starting from the random sampling of 10kb loci, 10 times to ensure that the topology was consistent across multiple runs. We then computed a majority rule consensus using all 1000 bootstrap replicates (10runs*100boostrap=1000 trees) to obtain support values (Figure 1a).

The resulting tree was identical to the tree in ref. 2,3, with 1000 bootstrap support for each node, except the node of Tibetan/North Chinese wild boar that had a support of 470 (Figure 1a).

We computed the divergence time for *Suinae* (African *Suidae* [Warthog]/*Sus*) based on evolutionary time estimates by Li *et. al.* for Eurasian wild boars. Assuming that Duroc and Tibetan pigs diverged 6.8Mya as in Li. *et. al.*, we can compute a rate of substitution per million years using our estimate of branch length inferred via ML (Figure 1a). This rate can then be utilized to compute a divergence time for the Suinae at roughly 26Mya (50-18Mya).

3. ***Demographic analysis***

We performed a PSMC[6] analysis to reconstruct the demographic history of European, Chinese and Tibetan Wild boars using genotype calls from GATK. For PSMC, we used the following parameters: T max = 20; n = 64 ('4+50*1+4+6'). For plotting the results we used g=5 and a rate of 2.5x10-8 mutations per generation as in ref 2,3.

	Total number of reads	Proportion of read mapped	Proportion of reads properly paired	Average depth of coverage
Tibetan Wild Boar library SRR652444	328,151,410	82%	76%	11.5x

Table S1: Alignment statistics for Tibetan Wild boar short-read library SRR652444.

References:

1. Ponstingl H.: SMALT [http://www.sanger.ac.uk/resources/software/smalt/]
2. Frantz, L.A.F. *et al.* Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome biology* **14**, R107 (2013).
3. Groenen M.A.M. *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398 (2012).
4. McKenna A. *et. al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
5. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22** 2688-90 (2006).
6. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-6 (2011).