Published in final edited form as:

Nat Genet. 2015 March ; 47(3): 188–189. doi:10.1038/ng.3197.

Evolution of Tibetan Wild Boars

Laurent A.F. Frantz^{1,5,*}, Ole Madsen¹, Hendrik-Jan Megens¹, Joshua G. Schraiber^{2,3}, Yogesh Paudel¹, Mirte Bosse¹, Richard P.M.A. Crooijmans¹, Greger Larson^{4,5}, and Martien A.M. Groenen^{1,*}

¹Animal Breeding and Genomics Group, Wageningen University, Wageningen 6700AH, The Netherlands ²Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA ³Department of Genome Sciences, University of Washington, Seattle, WA 98195-5065, USA ⁴Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham, DH1 3LE, UK

To the Editor:

The analysis presented by Li *et al.*¹ in their report of the genome sequence of the Tibetan wild boar provides interesting insights into the genetic architecture of high altitude adaptation in this species. However, despite the large volume of novel data, we found shortcomings in several parts of the study, suggesting that some specific findings presented by Li *et al.* result from over-interpretation. In addition, several of their conclusions contradict those reported in previous analyses^{2–5}.

More specifically, the authors infer that Tibetan wild boars and Duroc breeds (*S. scrofa* Ssc10.2 reference genome) diverged during the Miocene, ~6.8 million years ago (Mya). This estimated date is nearly 10 times more ancient than the recently reported split between Asian and European wild boar (2–0.8Mya)^{2,4}. In addition, previous studies^{2,3} estimated the divergence time between *S. scrofa* and other *Sus* species from Island Southeast Asia (outgroups in Figure 2b/e in ref 1) to be 5.3–1.3 Mya. Li *et al.*¹ do not describe the details of the molecular clock analysis, other than stating that PAML (MCMCTREE)⁶ was used with three molecular clock based calibrations (as opposed to fossil calibrations), and they neither specified which nodes were calibrated nor did they include the uncertainties of these calibrations in their age priors. Moreover, we believe that the tree used for the molecular clock analysis is too sparsely sampled to be informative for the Duroc-Tibetan split time. Indeed, different mutation rates are expected for the deep internal branches, separating mammalian orders, compared to the short branches separating the two sub-species of *S. scrofa*^{2,4,7}. Hence, estimating a subspecific split using rates estimated from the divergence

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

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

Author contributions:

L.A.F.F. and M.A.M.G designed study with input from H-J.M., O.M. and G.L. L.A.F.F analyzed the data with help from H-J.M., Y.P., M.B. and R.P.M.A.C. L.A.F.F wrote the manuscript with input from all authors.

Competing financial interests:

The authors declare no competing financial interests.

of mammalian orders is nearly certain to bias the estimate leading to an incorrect conclusion. We therefore believe that this analysis is likely to be influenced both by prior age misspecification and biased taxon sampling, resulting in a gross overestimation of the divergence time.

Furthermore, the authors do not take into account the wild boar paleontological literature that provides specific contradictory examples to their conclusions (Figure 1a) and would have provided useful calibrations (see Additional File 6 in ref 2). To illustrate our concerns we constructed a phylogenetic tree using data from a Tibetan wild boar¹ together with seven other *Sus* samples (Figure 1a) that were used in two previous studies^{2,4} and the study of Li *et al.*¹ (Supplementary Note). We further annotated the tree with well known fossils, and our own time estimates^{2,3,8} (Figure 1a). Our results demonstrate that Tibetan wild boar clusters together with Chinese wild boar (also suggested by the Li *et al.* ancestry and phylogenetic analyses). Based on our molecular clock analyses^{2,3}, we conclude that Tibetan and European wild boars are not different species, but instead are closely related sub-species that diverged during the Pleistocene. In addition, we show that the evolutionary time scale proposed by Li *et al.*¹ for the genus *Sus* implies a significantly older speciation timeframe contradicting the fossil record (Figure 1a). Furthermore, according to our branch length estimates (Supplementary Methods), a divergence time for European and Asian wild boars of 6.8 Mya¹ (confidence interval 12.9-2.4), would imply that African and Eurasian Suinae diverged roughly 26 Mya (confidence interval 50-13 Mya), an estimate that bears no correspondence with the fossil time-frame for Suinae⁸. The erroneous time inference presented by Li *et al.* contributed to their generally implausible results.

Another point of concern is the results of their demographic analysis⁹, pictured in Figure 2e in Li *et al.*¹ which highlights the low coverage of the dataset and the shortcomings of a next-generation-sequence (NGS) based genome assembly. The authors present the demographic history as inferred both from boars re-sequenced and aligned to their assembly (green and pink lines in Figure 2e in ref 1) as well as previously sequenced boar genomes and aligned to the high quality draft reference genome⁴ (Ssc10.2, blue and black lines Figure 2e in ref 1). Surprisingly, the results for Chinese wild boar aligned to Ssc10.2 and to the *de novo* assembly significantly disagree (*e.g.* South China with Southwest China in Figure 2e in ref 1.). This lack of correspondence is worrying since these individuals have similar ancestry and are very closely related (see Figure 2b in ref 1; Figure 1a). The authors do not address this discrepancy but instead draw the conclusion, “to our knowledge this is the first study supporting the refuge theory on the basis of demographic history revealed by genome-wide analysis”. We believe that this result does not reflect a realistic population size for wild boar in Tibet during the Pleistocene, but instead illustrates the lack of resolution of *de novo* assembled genomes for demographic estimations in combination with underestimation of true heterozygosity due to low coverage re-sequencing data (5x coverage for individuals re-sequenced in this study versus 10x coverage for previously published re-sequenced individuals). To support our claims, we re-analyzed Li *et al.*’s data by aligning short-read sequences from a single Tibetan wild boar that was used for *de novo* assembly (over 10x coverage; Supplementary Methods) to the Ssc10.2 reference genome and conducted a similar PSMC analysis to the one carried out by Li *et al.*¹ Our results demonstrate that

Tibetan and South Chinese wild boar have similar demographic histories (Figure 1b). In particular, we show that Tibetan wild boar underwent population size fluctuations during the Pleistocene, contrary to the conclusions of Li *et al.* This analysis shows that the claim that Tibetan wild boar experienced no demographic fluctuations due to the presence of refugia in Tibet is incorrect.

In the light of the Assemblathon 2.0¹⁰ we believe that it is important not to over-interpret data from NGS short-read *de novo* assemblies, especially when comparing these to an assembly built from Sanger reads combined with a high-resolution linkage map. Here we show that the over-confidence of the authors in their *de novo* assembly and divergence time estimates misled their interpretation of the evolutionary history of Tibetan wild boars. Furthermore, we believe that some results presented in the study by Li *et al.*¹, in particular large scale gene family expansion and/or contraction since the European-Tibetan wild-boar split are dubious. The interpretation of these results, as being the result of a genuine biological signal rather than assembly artifact, may have been misled by the gross over-estimation of evolutionary time frame of *S. scrofa*.

We suggest that the experimental design of studies describing re-sequencing of closely related species or sub-species needs to be different from the analysis conducted on newly sequenced genomes. For example, a more relevant analysis would have been to compare the “wild *Sus* species” and Warthog to the genome of the Tibetan wild boar rather than comparing within *S. scrofa* evolutionary changes with evolutionary processes that occurred at the scale of mammalian evolution. With the power of multiple complete genomes comes the responsibility to interpret them in light of both the potential pitfalls of NGS and the published genomic and paleontological evidence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Li M, et al. Nature genetics. 2013; 45:1431–1438. [PubMed: 24162736]
2. Frantz LAF, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. Genome biology. 2013; 14:R107. [PubMed: 24070215]
3. Gongora J, et al. Zoologica Scripta. 2011; 40:327–335.
4. Groenen MAM, et al. Nature. 2012; 491:393–398. [PubMed: 23151582]
5. Paudel Y, et al. BMC genomics. 2013; 14:449. [PubMed: 23829399]
6. Yang Z. Molecular biology and evolution. 2007; 24:1586–91. [PubMed: 17483113]
7. Ho SYW, Larson G. Trends in Genetics. 2006; 22:79–83. [PubMed: 16356585]
8. Orliac MJ, Pierre-Olivier A, Ducrocq S. Zoologica Scripta. 2010; 39:315–330.
9. Li H, Durbin R. Nature. 2011; 475:493–6. [PubMed: 21753753]
10. Bradnam KR, et al. GigaScience. 2013; 2:10. [PubMed: 23870653]

Figure 1. Evolutionary history of *Sus* species

a. Maximum likelihood phylogenetic tree for *Suinae* (see Supplementary Note). All nodes, besides node of Tibetan/North Chinese wild boar (support 47%) are supported by 100% bootstrap replicates. Nodes label represent time estimates in millions of years from ref. 2,3 and 1 respectively; *indicate divergence times that were converted using branch lengths estimated in this study and the times reported by Li *et al.*¹ (see Supplementary Note). Grey

dots represent well-known fossils, with time in millions of years. b. Demographic history of Eurasian wild boar (see Supplementary Methods).