Evolution of Tibetan wild boars

© 2015 Nature America, Inc. All rights reserved.

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 overinterpretation of the data. In addition, several of their conclusions contradict those reported in previous analyses[2–5].

More specifically, the authors infer that Tibetan wild boar and Duroc breeds (*Sus scrofa* Ssc10.2 reference genome) diverged during the Miocene, ~6.8 million years ago. This estimated date is nearly ten times more ancient than the recently reported split between Asian and European wild boars (0.8–2 million years ago)[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 Fig. 2b,e of ref. 1) to be 1.3–5.3 million years ago. 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-based calibrations), and they neither specify 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 was too sparsely sampled to be informative for the Duroc (*S. scrofa* reference genome) and Tibetan split time. Indeed, different mutation rates are expected for the deep internal branches separating mammalian orders and the short branches separating the two subspecies of *S. scrofa*[2,4,7]. Hence, estimating a subspecies split time using rates estimated from the divergence of mammalian orders is nearly certain to bias the estimate, leading to an incorrect conclusion. We therefore believe that this analysis is likely to have been influenced both by prior age misspecification and biased taxon sampling,

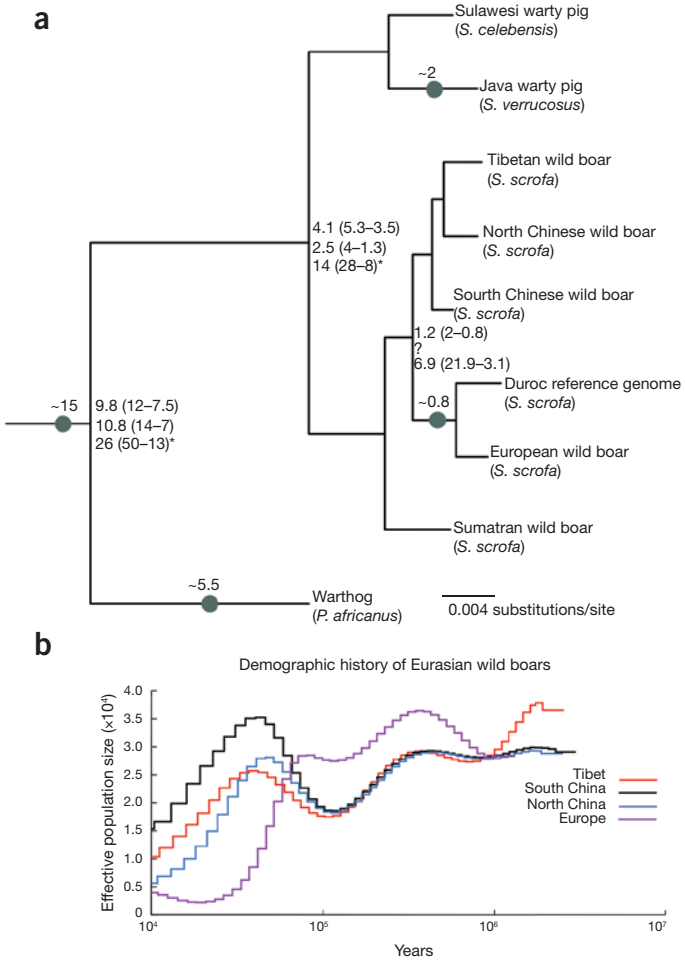

Figure 1 Evolutionary history of *Sus* species. (**a**) Maximum-likelihood phylogenetic tree for Suinae (**Supplementary Note**). All nodes besides the node for Tibetan or northern Chinese wild boars (support of 47%) are supported by 100% bootstrap replicates. Node labels represent time estimates in millions of years from refs. 2, 3 and 1, respectively; an asterisk indicates divergence times that were converted using branch lengths estimated in this study and the times reported by Li *et al.*[1] (**Supplementary Note**). Gray dots represent well-known fossils, with time in millions of years. (**b**) Demographic history of Eurasian wild boar (**Supplementary Note**). Generation time (g) = 5, mutation rate (μ) = 2.5×10^{-8}.

resulting in a gross overestimation of the subspecies divergence time.

Furthermore, the authors do not take into account the suid paleontological literature that provides specific examples contradictory to their conclusions (**Fig. 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 boar[1] together with seven other *Sus* samples (**Fig. 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] (**Fig. 1a**). Our results demonstrate that

Tibetan wild boar clusters together with Chinese wild boar (as also suggested by the ancestry and phylogenetic analyses of Li *et al.*). On the basis of our molecular clock analyses[2,3], we conclude that Tibetan and European wild boars are not different species but instead are closely related subspecies that diverged during the Pleistocene. In addition, we show that the evolutionary time scale proposed by Li *et al.*[1] for the *Sus* genus implies a significantly older speciation timeframe that contradicts the fossil record (**Fig. 1a**). Furthermore, according to our branch length estimates (**Supplementary Note**), a divergence time for European and Asian wild boars of 6.8 million years ago[1] (confidence interval of 2.4–12.9 million years ago) would imply that African and Eurasian Suinae diverged roughly 26 million years ago (confidence interval of 13–50 million years ago), an estimate that bears no correspondence with the fossil timeframe for Suinae[8]. The erroneous time inference presented by Li *et al.* contributes to their generally implausible results.

Another point of concern is the results of their demographic analysis[9], shown in Figure 2e of Li *et al.*[1], which highlight the low coverage of the data set and the shortcomings of a next-generation sequence–based genome assembly. The authors present the demographic history as inferred both from boar genomes resequenced and aligned to their assembly (green and pink lines in Fig. 2e in ref. 1) and previously sequenced boar genomes aligned to the high-quality draft reference genome[4] (Ssc10.2; blue and black lines in Fig. 2e in ref. 1). Surprisingly, the results show that the alignments of Chinese wild boar to Ssc10.2 and the *de novo* assembly significantly disagree (compare South China with Southwest China in Fig. 2e in ref. 1). This lack of correspondence is worrying as these individuals have similar ancestry and are very closely related (**Fig. 1a**; see also Fig. 2b in ref. 1). 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 resequencing data (5× coverage for individuals resequenced in Li *et al.*[1] versus 10× coverage for previously published resequenced individuals). To support our claims, we reanalyzed the data of Li *et al.* by aligning short-read sequences from a single Tibetan wild boar that was used for *de novo* assembly (over 10× coverage; **Supplementary Note**) to the Ssc10.2 reference genome and conducted a similar pairwise sequentially Markovian coalescence (PSMC) analysis to the one carried out by Li *et al.*[1]. Our results demonstrate that Tibetan and southern Chinese wild boars have similar demographic histories (**Fig. 1b**). In particular, we show that Tibetan wild boar underwent fluctuations in population size 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 owing to the presence of refugia in Tibet is incorrect.

In light of the Assemblathon 2.0 (ref. 10), we believe that it is important not to overinterpret data from *de novo* assemblies of next-generation short-read sequencing data, especially when comparing these to an assembly built from Sanger sequencing reads combined with a high-resolution linkage map. Here we show that the overconfidence 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 due to a genuine biological signal rather than assembly artifact might have been misled by the gross overestimation of the evolutionary timeframe of *S. scrofa*.

We suggest that the experimental design of studies describing the resequencing of closely related species or subspecies needs to be different from the analysis conducted on newly sequenced genomes. For example, a more relevant analysis would have been to compare the genomes of the 'wild' *Sus* species and warthog[2,4] to the genome of the Tibetan wild boar rather than comparing evolutionary processes that occurred during the last million years (Asian-European pig split) with processes that occurred since the common ancestor of human and pigs, roughly 100 million years ago. With the power of multiple complete genomes comes the responsibility to interpret them in light of both the potential pitfalls of next-generation sequencing and the published genomic and paleontological evidence.

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

AUTHOR CONTRIBUTIONS
L.A.F.F. and M.A.M.G. designed the 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., R.P.M.A.C. and J.G.S. L.A.F.F. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

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] & 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, Berkeley, California, USA. [3]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [4]Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, Durham, UK. [5]Present address: Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford, UK.
e-mail: laurent.frantz@arch.ox.ac.uk or martien.groenen@wur.nl

1. Li, M. *et al. Nat. Genet.* **45**, 1431–1438 (2013).
2. Frantz, L.A.F. *et al. Genome Biol.* **14**, R107 (2013).
3. Gongora, J. *et al. Zool. Scr.* **40**, 327–335 (2011).
4. Groenen, M.A.M. *et al. Nature* **491**, 393–398 (2012).
5. Paudel, Y. *et al. BMC Genomics* **14**, 449 (2013).
6. Yang, Z. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
7. Ho, S.Y.W. & Larson, G. *Trends Genet.* **22**, 79–83 (2006).
8. Orliac, M.J., Pierre-Olivier, A. & Ducrocq, S. *Zool. Scr.* **39**, 315–330 (2010).
9. Li, H. & Durbin, R. *Nature* **475**, 493–496 (2011).
10. Bradnam, K.R. *et al. Gigascience* **2**, 10 (2013).

Li *et al.* reply:

Frantz *et al.* reanalyzed our data[1] by aligning a small part of our short-read sequences from a Tibetan wild boar that were used for *de novo* assembly (33.2 Gb of 319 Gb, or ~10%) to the reference genome for the European domesticated Duroc pig breed (*Sus scrofa* Ssc10.2)[2]. As a result, they propose alternative interpretations and question two of our findings. They also suggest that a resequencing strategy should be applied for the comparison of Tibetan wild boar and Duroc pig, rather than for the comparison of the two *de novo* assemblies as we reported.

First, Frantz *et al.* claim that our proposed time of divergence between Duroc pig and Tibetan wild boar is an overestimation that can be attributed to the lack of appropriate calibration points and insufficient sampling in the *Sus* lineage. Using three fossil-based calibration points and resequencing data from eight *Sus* samples (including our Tibetan wild boar) and an outgroup species,

© 2015 Nature America, Inc. All rights reserved.
npg