






# Addendum: Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion

Langqing Liu , Mirte Bosse , Hendrik-Jan Megens, Laurent A. F. Frantz, Young-Lim Lee, Evan K. Irving-Pease , Goutam Narayan , Martien A. M. Groenen  & Ole Madsen

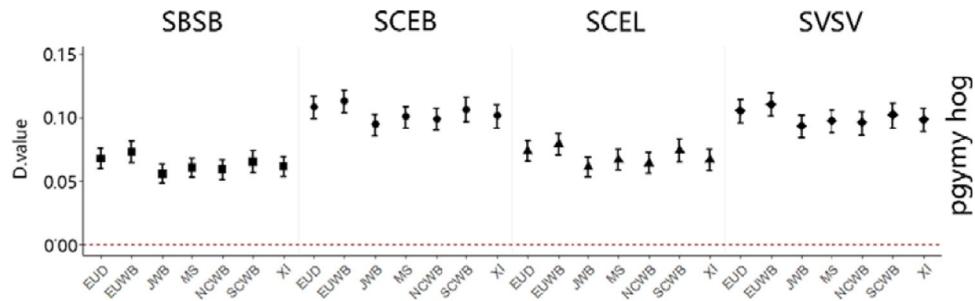
Correction to: *Nature Communications* <https://doi.org/10.1038/s41467-019-10017-2>, published online 30 April 2019.

In the original article, we analyzed 38 genomes from pygmy hog and related suid species. Our analysis not only identified a signal of introgression between *Sus scrofa* and pygmy hog but also reinforced the idea that there was gene flow between *Sus scrofa* and an extinct (ghost) *Suidae* lineage<sup>1</sup>. In our original analysis, however, for the *D*-statistics equation in Admixtools<sup>2</sup>, we mistakenly interpreted as  $D = \frac{ABBA-BABA}{ABBA+BABA}$ , while it was in fact  $D = \frac{BABA-ABBA}{ABBA+BABA}$ . This led to an inversion of the direction of admixture shown in Fig. 2b in the original version of the article. To remain consistent with the rest of the analyses in the article, we rectified the formula of *D*-statistics as  $D = \frac{ABBA-BABA}{ABBA+BABA}$  and updated Fig. 2b (Fig. 1; this addendum). The results after correcting the formula indicate that in fact there is an excess of shared derived alleles between the pygmy hog and Island of Southeast Asian (ISEA) *Sus* (ABBA) and not between the pygmy hog and *Sus scrofa* (BABA). The excess of ABBA can be the result of admixture either between pygmy hog and ISEA *Sus* or between *Sus scrofa* and the archaic ghost lineage or both. Methods to detect hybridization such as Patterson's *D*, *fd*, or *Twisst*<sup>3–5</sup>, however, are inadequate to distinguish gene-flow events between P1 and P3 (Fig. 2, orange arrow) from those between P2 and an archaic ghost lineage basal to P3 (Fig. 2, blue arrow).

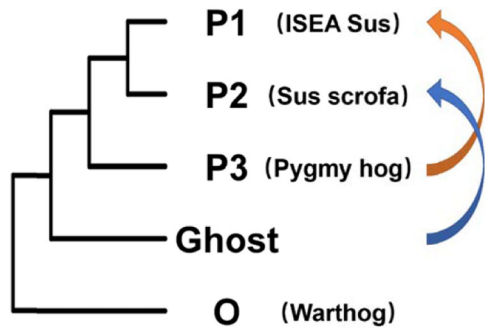
To address this issue an alternative method was used, based on relative nucleotide distance, as implemented in  $RND_{min}$ <sup>6</sup> to calculate the relative node depth between each taxon. To do so, we first used BEAGLE v4.1<sup>7</sup> to perform haplotype phasing on genotypes of all individuals (excluding *Babyrussa babyrussa*, *Potamochoerus larvatus*, and *Potamochoerus porcus*) with default parameters. Then we calculated  $RND_{min}$  in 100-kb sliding windows along autosomes with <50% missing sites using PopGenome<sup>8</sup>. Comparisons were carried out between pygmy hog, ISEA *Sus*, and *Sus scrofa*, using the genome of a warhog as an outgroup. This shows that, for the genomic regions supporting topoA (Fig. 3a), the distribution of  $RND_{min}$  computed between the pygmy hog and *Sus scrofa* is significantly lower (16.2%, Welch's *t* test *p* value <1.8e–14) than  $RND_{min}$  between pygmy hog and ISEA *Sus*, indicating that the first two species share more similarities. We also note that this pattern is more apparent in regions supporting topoA than in the overall genomic background (gray dots in Fig. 3a). Altogether, this indicates that the average genomic distance between the pygmy hog and *S. scrofa* is smaller than that between the pygmy hog and the ISEA *Sus* clade, which supports the existence of gene flow between pygmy hog and *S. scrofa*.

If the result of our *D*-statistics were affected by an admixture between pygmy hog and ISEA *Sus*, we would expect that, in regions that display topoB (Fig. 3b), the  $RND_{min}$  computed between the pygmy hog and ISEA *Sus* should be lower than in the overall genomic background. Our analysis, however, shows that this is not the case. In fact, we found these distributions to be very similar (Fig. 3b), especially when compared to the result shown in Fig. 3a. In addition, we found the distance between ISEA *Sus* and *Sus scrofa* to be higher (23.2%, Welch's *t* test *p* value <2.2e–16) than the overall genomic background in the regions that show topoB, suggesting that topoB was caused by the admixture with an unsampled taxon.

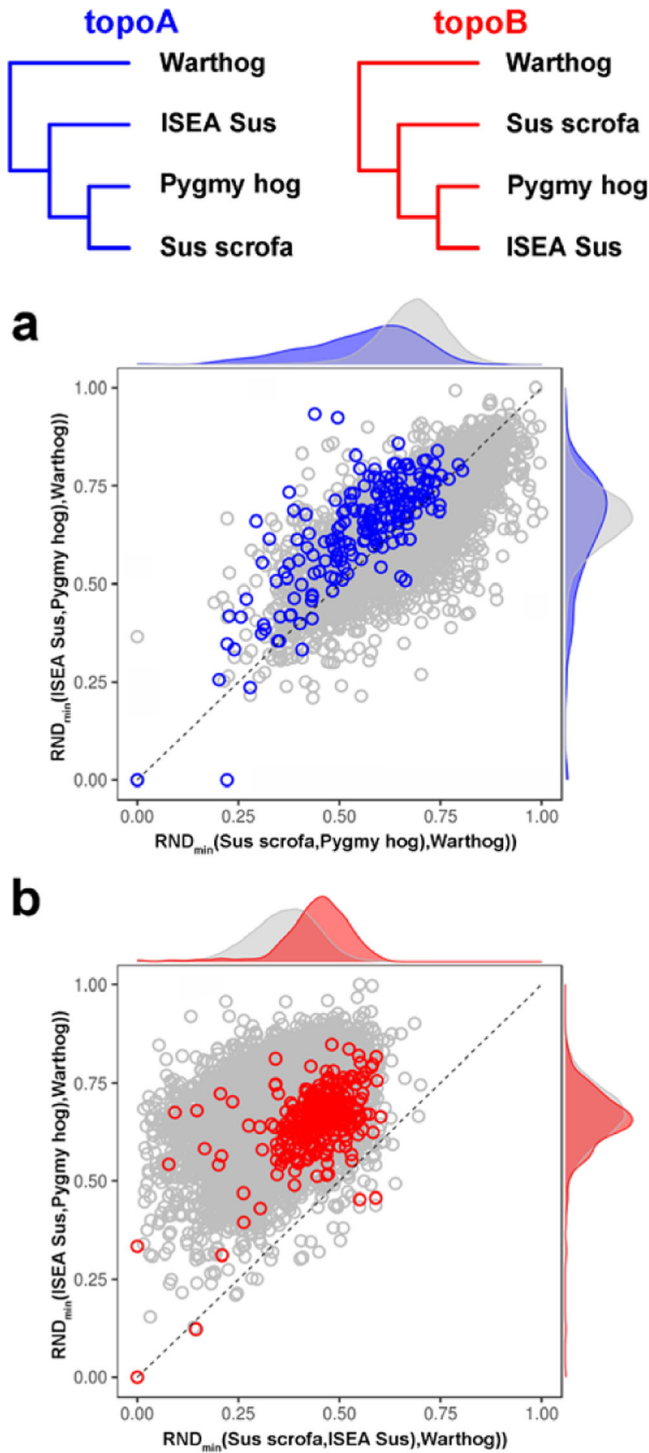
Altogether, our analyses indicate that the findings and interpretations presented in the original article are correct. First, there was no admixture between ISEA *Sus* and pygmy hog. Instead, the combined results of the *D*-statistics and  $RND_{min}$  are more consistent with a ghost admixture between a distantly related taxon and *S. scrofa*, and this inflated the number of shared derived alleles between ISEA *Sus* and the pygmy hog. Second, the shorter distance between *Sus scrofa* and pygmy hog suggests that there was also admixture between these two species.



**Fig. 1 A diagram depicting the excess derived allele sharing when comparing sister taxa and outgroups.** Each column contains the fraction of excess allele sharing by a taxon (up/down) with the pygmy hog/outgroup compared to its sister taxon (up/down). We computed  $D$  statistics of the form  $D(X, Y, \text{Pygmy hog, warthog})$ . Error bars correspond to three standard errors. (SBSB *Sus barbatus*, SCEB *Sus cebrifons*, SCEL *Sus celebensis*, SVSV *Sus verrucosus*, EUD European domesticated pig, EUWB European wild boar, JWB Japanese wild boar, MS Meishan, NCWB Northern China wild boar, SCWB Southern China wild boar, XI Xiang).



**Fig. 2 Schematic representation of the potential gene-flow scenarios discussed in the text.** Two-way arrows pointing the two hybridized species.



**Fig. 3 Phylogenetic tree indicated the alternative topologies (topoA and topoB) to the main species tree.** Scatter and density plots below show the distribution of  $RND_{min}$  between different comparisons. **a** Blue circles represent  $RND_{min}$  distribution within windows supporting topoA. **b** Red circles represent  $RND_{min}$  distribution within windows supporting topoB. Gray circles represent  $RND_{min}$  distribution of the 100-kb windows among all autosomes. Populations used to calculate  $RND_{min}$  is shown on axis labels. Warthog was used as outgroup.

Published online: 03 December 2020

## References

1. Ai, H. et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
2. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
3. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
4. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
5. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
6. Rosenzweig, B. K., Pease, J. B., Besansky, N. J. & Hahn, M. W. Powerful methods for detecting introgressed regions from population genomic data. *Mol. Ecol.* **25**, 2387–2397 (2016).
7. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
8. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020