African bush pigs exhibit porous species boundaries and appeared in Madagascar concurrently with human arrival

Authors

Renzo F. Balboa[1#], Laura D. Bertola[1#], Anna Brüniche-Olsen[1#], Malthe Sebro Rasmussen[1], Xiaodong Liu[1], Guillaume Besnard[2], Jordi Salmona[2], Cindy G. Santander[1], Shixu He[1], Dietmar Zinner[3,4,5], Miguel Pedrono[6], Vincent Muwanika[7], Charles Masembe[8], Mikkel Schubert[1], Josiah Kuja[1], Liam Quinn[1], Genís Garcia-Erill[1], Rianja Rakotoarivony[6], Margarida Henrique[9], Long Lin[1], Xi Wang[1], Michael P. Heaton[10], Timothy P. L. Smith[10], Kristian Hanghøj[1], Mikkel-Holger S. Sinding[1], Anagaw Atickem[11], Lounès Chikhi[2,9], Christian Roos[12], Philippe Gaubert[2], Hans R. Siegismund[1], Ida Moltke[1]*, Anders Albrechtsen[1]* and Rasmus Heller[1]*

[1] Department of Biology, University of Copenhagen, Copenhagen, Denmark

[2] Laboratoire Evolution et Diversité Biologique (EDB), UMR 5174, CNRS, IRD, Université Toulouse Paul Sabatier, 31062 Toulouse, France

[3] Cognitive Ecology Laboratory, German Primate Center, Leibniz Institute for Primate Research; 37077 Göttingen, Germany

[4] Department of Primate Cognition, Georg-August-Universität Göttingen; 37077 Göttingen, Germany

[5] Leibniz Science Campus Primate Cognition; 37077 Göttingen, Germany

[6] UMR ASTRE, CIRAD, Campus International de Baillarguet, Montpellier, France

[7] College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda

[8] College of Natural Sciences, Makerere University, Kampala, Uganda

[9] Instituto Gulbenkian de Ciência, Oeiras, Portugal

[10] USDA, ARS, US Meat Animal Research Center, Clay Center, Nebraska, USA

[11] Department of Zoological Sciences, Addis Ababa University, PO Box 1176, Addis Ababa, Ethiopia

[12] Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany

[#] These authors contributed equally

* Corresponding/senior author

Abstract

Several African mammals exhibit a phylogeographic pattern where closely related taxa are split between West/Central and East/Southern Africa, but their evolutionary relationships and histories remain controversial. Bushpigs (*Potamochoerus larvatus*) and red river hogs (*P. porcus*) are recognised as separate species due to morphological distinctions, a perceived lack of interbreeding at contact, and putatively old divergence times, but historically, they were considered conspecific. Moreover, the presence of Malagasy bushpigs as the sole large terrestrial mammal shared with the African mainland raises intriguing questions about its origin and arrival in Madagascar. Analyses of 67 whole genomes revealed a genetic continuum between the two species, with putative signatures of historical gene flow, variable F_{ST} values, and a recent divergence time (<500,000 years). Thus, our study challenges key arguments for splitting *Potamochoerus* into two species and suggests their speciation might be incomplete. Our findings also indicate that Malagasy bushpigs diverged from southern African populations and underwent a limited bottleneck 1,000-5,000 years ago, concurrent with human arrival in Madagascar. These results shed new light on the evolutionary history of an iconic and widespread African genus and provide insight into the longstanding biogeographic puzzle surrounding the bushpig's presence in Madagascar.

Keywords: whole genomes, genomics, population genetics, metapopulations, introgression, hybridisation, speciation, Suidae.

Introduction

The African Suidae lineage contains six recognised extant species: common warthog (*Phacochoerus africanus*), desert warthog (*Ph. aethiopicus*), giant forest hog (*Hylochoerus meinertzhageni*), wild boar (*Sus scrofa)*, red river hog (*Potamochoerus porcus*) and bushpig (*P. larvatus*) [1,2]. There are several unresolved aspects of the evolutionary history of African pigs, including a controversial timeline for their divergence which stems from molecular estimates that predate fossil records by millions of years, and the unresolved role of gene flow between lineages [3,4]. The two members of the genus *Potamochoerus* — red river hog and bushpig — were historically considered conspecific, despite considerable morphological differences [5,6]. They occur parapatrically in Western/Central (W/C) Africa and Eastern/Southern (E/S) Africa with some populations possibly having abutting or slightly overlapping ranges [2] (Fig. 1a). Based primarily on morphological differences and a lack of evidence that these taxa hybridise at contact, Grubb proposed the currently accepted nomenclature, regarding them as two distinct species [7,8].

The distribution of the two *Potamochoerus* species is similar to that found in several other African mammals that have ecologically comparable sister (sub)species pairs. The W/C and E/S divide has been highlighted as one of the most important biogeographic patterns in Africa, and is potentially connected to the initial divergence between hominins and apes [9], even if at a different time scale. This evolutionary divergence into W/C and E/S lineages occurred relatively recently for some mammalian taxa such as the African buffalo (*Syncerus caffer*) [10] and the lion (*Panthera leo*) [11] leading to subspeciation, whereas in other taxa, an older split led to full speciation, e.g. in African elephants (*Loxodonta* sp.)[12] and baboons (*Papio* sp.)[13]. For all species mentioned above, a hybrid zone has been identified where the ranges of diverged lineages overlap [8]. Although possible hybridisation between the two *Potamochoerus* species has been suggested [7], the evolutionary connection and the geographic context of a likely suture zone are still poorly understood [5,14]. Recent range contractions limit the overlap of the two species ranges to Uganda and Democratic Republic of Congo (DR Congo). However, South Sudan and possibly Ethiopia were part of a suture zone in the recent past, when the red river hog range extended further towards the East (Fig. 1a) [5]. The evolutionary processes occurring in these suture zones, found recurrently across many taxa, e.g. in western Uganda [10,15], are of particular interest for understanding speciation and the phylogeography of African mammals in general.

Bushpig populations on Madagascar provide an interesting case of possible human-mediated range expansion. The bushpig represents a biogeographic anomaly in being the only large,

wild terrestrial mammal to be shared between the African continent and the island of Madagascar [16]. These land masses separated about 150 million years ago, leading to a largely divergent fauna and flora [17,18]. For some Malagasy taxa, such as lemurs, it has long been debated whether colonisation of Madagascar could have taken place through island hopping or temporal land bridges [19]. It is now commonly accepted that some of these taxa arrived on Madagascar by rafting on floats of vegetation, and that successful colonisation events and subsequent radiation led to the diversity seen today [20,21]. For bushpigs, it has been proposed that the most plausible explanation is that they were introduced to Madagascar by humans, possibly through the Comoros Islands [22,23]; however, this has not been conclusively verified. Humans are believed to have been present in Madagascar no earlier than 11,000 years ago [24], with some authors claiming that there is no proof of human presence older than 2,000 years [25]. Nevertheless, most authors agree that there were no significant numbers of humans until 1,000-1,500 years ago with the arrival of populations from South-Eastern Africa (Bantu speakers) and South-East Asia (Austronesian speakers) [24,26,27]. Radiocarbon dating of archaeological remains suggests that bushpigs, as well as zebu, sheep and goats, were established in southwest Madagascar between 700-1,200 years ago; however, this estimate may be influenced by the scarce data available for Malagasy bushpigs [28]. To our knowledge, there is only one study which attempted to estimate the arrival of bushpigs on Madagascar based on genetic data; this study suggested a split time of 480 kya based on mitochondrial DNA (mtDNA) divergence times, which is not in line with a proposed human-mediated introduction to the island [29]. In addition to the time of arrival, the source population for Malgasy bushpigs is still unknown, where despite detailed morphological studies, these have been unable to conclusively resolve their mainland origin [8,30]. The existing genetic data tentatively suggest an origin from Central Southern Africa [29]. If bushpigs were indeed introduced to Madagascar by humans, it presents another suite of questions as there is no archaeological or other evidence of domestication of bushpigs ever occurring despite them being an important protein source for many rural communities [31]. For example, the transportation of such a large non-domesticated mammal over the wide (> 400 km) Mozambique channel remains an unsolved mystery, and may provide an indirect indication that populations located on the south-eastern African coast mastered oceanic travel beyond fishing [29]. Alternatively, a much older divergence time could provide indirect proof of a very early African presence in Madagascar.

In this study, we present new data and population genomic analyses of 67 whole genomes from *Potamochoerus*, including 32 bushpigs from Madagascar. We investigate their population structure and genetic diversity, and infer gene flow between the two taxa. We also estimate the degree of evolutionary divergence between the bushpig and red river hog relative

to co-occurring species that represent incomplete or full speciation. Finally, we address the question of when and from where in Africa the bushpig colonised Madagascar, clarifying several details regarding the origin of Malagasy bushpigs. Our analyses present new insights that will improve our understanding of African biogeography, and help settle a major question regarding prehistoric human activities shaping biodiversity patterns in Africa.

Results

Sampling and filtering

Whole genome sequencing data were generated for 71 *Potamochoerus* samples across the two species' ranges, including 23 red river hogs and 48 bushpigs (3×-101×, mean ~12.8×; Fig. 1a; Supplementary Data 1). All samples were mapped to a chromosome-level common warthog reference genome, and rigorous site filtering applied to reduce downstream errors (see Methods; Supplementary Data 1). Two red river hog samples, from Cameroon and DR Congo, were excluded due to high sequencing error rates (Supplementary Fig. S1). Four samples, two from Equatorial Guinea (Eq Guinea) and two from Ethiopia, were deemed to originate from the same individual and were merged into one sample for their respective localities (Supplementary Fig. S2). A total of 13 samples were first degree relatives (parent-offspring or full siblings), of which 11 were from Madagascar and two were from Uganda. Depending on the specific requirements of the various downstream analyses, these samples were excluded. In summary, whole genome sequencing data from 67 pigs from 13 countries were analysed in this study, of which 54 were not closely related, including 18 that were sequenced at medium-high depth (≥14×; Fig. 1a; Supplementary Data 1). A summary of datasets, analyses and methods used is provided in Supplementary Data 1.

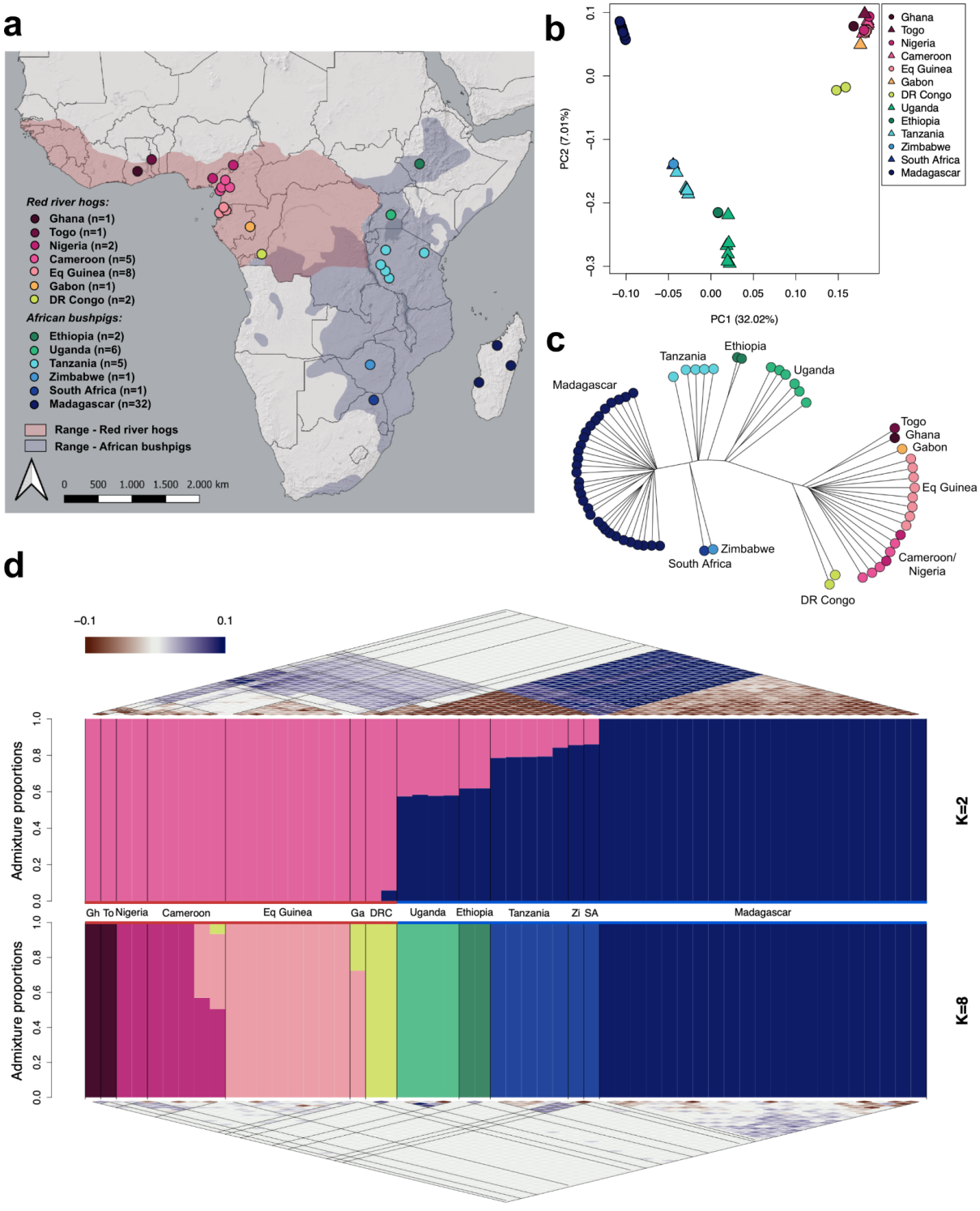

Figure 1. Sampling and population structure of red river hogs and bushpigs. a) Sampling map of all 67 pig individuals used within this study, coloured by country of origin. Ranges for red river hogs and bushpigs are shaded in red and blue, respectively [32,33]. **b)** Principal component analysis (PCA) for 67 pigs, showing the first two principal components, coloured by country. **c)** Unrooted neighbour-joining tree based on pairwise identity-by-state (IBS), coloured by country (n = 67). **d)** Inferred ancestry proportions for 54 unrelated samples using NGSadmix [34], assuming *K* = 2 (upper barplot) and *K* = 8 (bottom barplot). Coloured lines above and below population labels indicate species designations; red - red river hogs, blue - bushpigs. Pairwise correlations of residuals as estimated by evalAdmix [35] are shown above and below the respective NGSadmix barplots. Gh - Ghana, To - Togo, Ga - Gabon, DRC - Democratic Republic of Congo, Zi - Zimbabwe, SA - South Africa.

Localised population structure and no recent admixture between red river hogs and bushpigs

We first aimed to gain insights into the population structure of red river hogs and African bushpigs, specifically examining genetic differentiation between populations of both species [34]. Principal component analysis (PCA) revealed that the first two principal components exhibited a spatial distribution pattern reflecting the taxonomic and geographic origins of the sampled pigs; the red river hog samples clustered together, with only the Congo individuals being closer to the bushpigs than the other red river hogs and the Malagasy samples formed a separate cluster from the other bushpigs (Fig. 1b). A neighbour-joining tree using identity-by-state delineated a clear division between red river hogs and bushpigs, displaying a basal split between the two groups (Fig. 1c). The tree also revealed more localised population substructure, including the Malagasy samples forming a clade separate from the other bushpigs.

We next inferred ancestry proportions within both putative species to further explore population substructure. Assuming the number of ancestral populations was 2 (K = 2), the result largely aligned with the pattern observed in PC1-PC2 (Fig. 1d). Notably, we did not observe a clear separation of red river hogs and bushpigs at K = 2, even when excluding Madagascar samples (Supplementary Fig. S3). It is worth noting that evalAdmix [35] indicated unresolved substructure, suggesting that this pattern should not be interpreted as the result of admixture and these numbers as admixture proportions. We obtained a much better fit by assuming a higher number of ancestral populations (K = 8, Fig. 1d; Supplementary Fig. S4) and were able to assign most geographic locations to their own ancestral population. However, this analysis did not reveal evidence of recent gene flow between bushpigs and red river hogs.

Moderate differentiation and gene flow between red river hogs and bushpigs through Uganda

Genetic differentiation between all pairs of populations was assessed using Hudson's F_{ST} and D_{xy} between unrelated individuals (Fig. 2a) [36,37]. F_{ST} values generally correlate with geographic distance (Fig. 2a; Supplementary Fig. S5). Notably, Ethiopia and Madagascar exhibited higher F_{ST} values (0.345-0.581 and 0.168-0.546, respectively). Excluding these populations, F_{ST} values within red river hog populations ranged from 0.023-0.286, while F_{ST} values within bushpigs were between 0.044 and 0.175. When comparing across species, F_{ST} excluding Ethiopia was higher (0.232-0.546), though not markedly so when compared to the most differentiated population pairs in within-species comparisons. D_{xy} values exhibited a similar trend, displaying increased nucleotide diversity between species relative to within-species

196 comparisons (Fig. 2a). As with F_{ST}, D_{xy} also correlated with geographic distance with the
197 exception of Tanzania; this population had increased D_{xy} relative to other populations. Notably,
198 D_{xy} for Ethiopia was similar to those between other bushpig and red river hog populations,
199 suggesting that the high F_{ST} observed for Ethiopia was likely driven by lower within-population
200 diversity. In contrast, the Ugandan bushpig population exhibited a reduced D_{xy} relative to other
201 bushpig populations, suggesting potential gene flow with the red river hogs. In fact, the lowest
202 D_{xy} between species was between the Ugandan and Congolese populations, which were also
203 the two geographically closest.
204

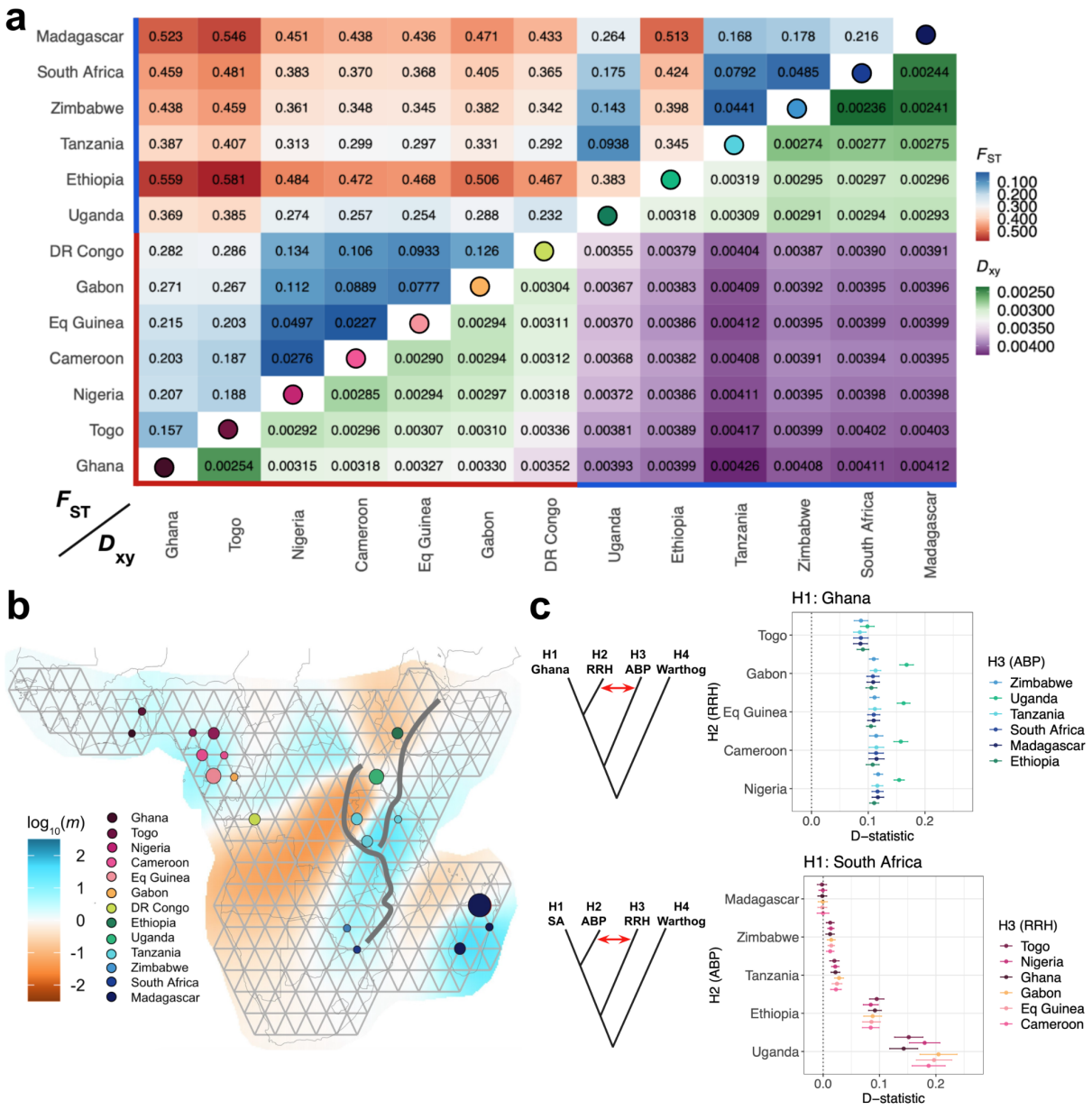

Figure 2. Genetic differentiation and gene flow between African bushpigs and red river hogs. a) Genetic differentiation as described by pairwise Hudson's F_{ST} [36] and D_{xy} [37] for 54 unrelated individuals, rounded to three significant figures. Circles on the diagonal correspond to populations as in b). Coloured lines above and next to population labels indicate species designations; red - red river hogs, blue - bushpigs. **b)** Estimated effective migration surfaces using EEMS [38]. Circles are coloured by country of origin. $\log_{10}(m)$ describes the effective migration rate relative to the overall migration rate across indicated regions. The East African Rift Valley is depicted by grey lines. **c)** D-statistics between populations using the common warthog as an outgroup, constructed as D(H1,H2,H3,Warthog). A significant non-zero positive value, as depicted by the red arrow in the graphic for each panel, provides evidence for gene flow between H3 and H2, relative to H1 (i.e. H2 is closer to H3 than H1) [39]. *Upper panel* - D-statistics testing for gene flow signals between African bushpigs (H3) and non-Ghana red river hogs (H2). *Lower panel* - D-statistics testing for gene flow signals between red river hogs (H3) and non-Malagasy bushpigs (H2). Error bars represent ± three standard errors from the estimated D-statistic. RRH - Red river hog; ABP - African bushpig; SA - South Africa.

Given the observed F_{ST} and D_{xy} values, we explored spatial patterns of gene flow between species (Fig. 2b) [38]. A general barrier through the Central African rainforest and following the East African Rift Valley was observed, separating W/C and E/S populations. Within each of the species ranges, connectivity was high, with the exception of Malagasy and non-Malagasy bushpigs where we observed a barrier across the Mozambique Channel, particularly with the northernmost non-Malagasy populations. We also observed a decrease in effective migration in Ethiopia. This is in contrast with Uganda where we observed weak gene flow barriers, suggesting a corridor of gene flow connectivity involving Uganda.

To detect potential gene flow patterns between the two species, we then tested for ancient admixture events. *f*-branch statistics revealed putative signals of gene flow between bushpigs and red river hogs, showing extensive gene flow involving Uganda (Supplementary Fig. S6). D-statistics were then used specifically to test whether there is increased allele sharing between red river hogs and bushpigs within the putative suture zone, compared to populations further from the suture zone (e.g. Ghana and Madagascar). Two tests were designed in order to test this hypothesis. We first set the westernmost population (Ghana) as H1, each of the red river hogs as H2 and each of the bushpigs as H3 (Fig. 2c; upper panel). This revealed that all red river hog populations showed signs of gene flow from African bushpigs, decreasing in signal from Central Africa to the westernmost red river hog populations, and with a particularly strong signal from Uganda. Similarly, to test for gene flow in the opposite direction, we performed similar tests with the easternmost bushpig population, Madagascar as H1, each of the remaining red river hog populations as H2, and each of the bushpig populations as H3 (Fig. 2c; lower panel). We observed a similar result, whereby we perceived a signal decrease towards more eastern and southern populations. Notably, this signal was particularly strong in Ethiopia and Uganda, suggesting substantial gene flow between red river hogs and these bushpig populations. These results suggest that there is or has been gene flow between the two taxa currently identified as species, and that the gradient of allele sharing between them is consistent with isolation by distance, where genetic similarity is strongest in populations from Central Africa. Additionally, these results could also be interpreted as a complex network of populations connected by genetic exchange, either recent or ancient.

Demographic histories and genetic diversity

Demographic histories of the surveyed populations were next explored (Fig. 3a) [40,41]. All PSMC curves overlapped from the most ancient past until ~500 kya, where we observed a stark difference in PSMC trajectories between red river hog and bushpig individuals. All red river hog populations first experienced a moderate increase (population expansion assuming

panmixia) followed by a more recent contraction ~50 kya. In contrast, bushpig individuals exhibited PSMC curves that followed three different trajectories: i) the populations in Tanzania, Zimbabwe, South Africa and Madagascar exhibited relatively constant (i.e. horizontal) curves until ~10 kya; ii) the Ugandan population showed a demographic history more similar to red river hogs than to the remaining bushpig populations, particularly between 100-500 kya and; iii) the Ethiopian population showed a history characterised by a declining and low PSMC curve ~200 kya. Given the results reported above, the unique demographic histories in Uganda and Ethiopia could be influenced by their geographic location as a place of introgression between the two taxa.

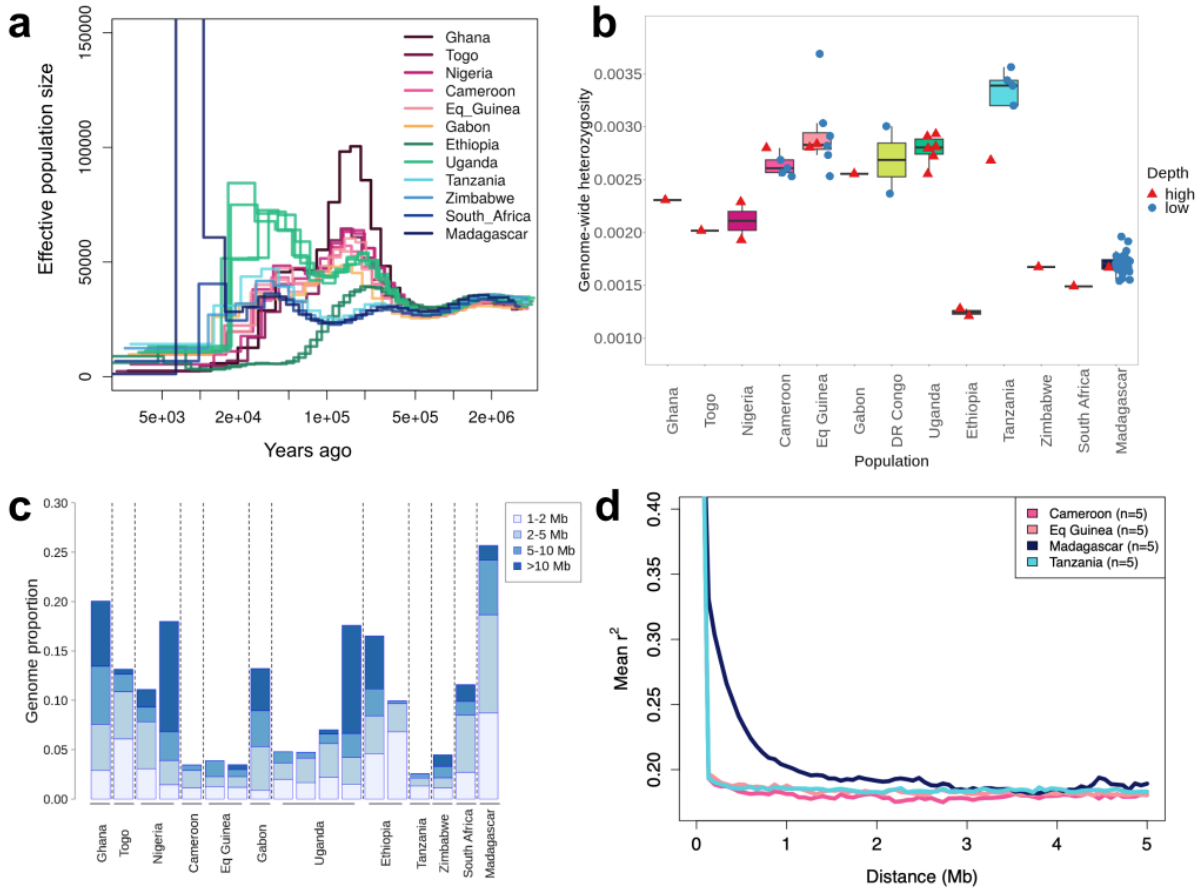

Figure 3. Genetic diversity of African pigs. a) Effective population sizes over time for 18 medium-high depth pig samples as estimated by PSMC, assuming a mutation rate of μ = 1.49e^{-8} per site per generation and a generation time of six years [4,42]. **b)** Genome-wide heterozygosity measurements described as the proportion of heterozygous sites per bp across each individual genome. Medium-high depth (n = 18; red triangles) and low-depth samples (n = 49; blue circles) are shown. **c)** Estimated genome-wide runs of homozygosity (ROH) proportions for 18 medium-high depth individuals. Each bar represents a single individual, grouped by their population. Proportions of differing ROH length intervals are shown as subdivisions within bars. **d)** Linkage-disequilibrium decay for populations with five or more samples, described as mean r^2 values for SNP pairs 0-5 Mb apart (n = 5 for each population).

Per-sample heterozygosity was next explored as a measure of genetic diversity, differing at both a species and population level (Fig. 3b). Heterozygosity was generally lower in bushpigs when compared to red river hogs, with the exception of Uganda and Tanzania which had similar heterozygosity levels to populations of DR Congo and Equatorial Guinea. The bushpig population in Ethiopia exhibited extremely low genetic diversity, one third of that of the highest, Tanzania (Fig. 3b). This was consistent with elevated F_{ST} values, reduced connectivity in EEMS and the low effective population size estimated by PSMC (Fig. 2a). Heterozygosity in Madagascar was also relatively low, but similar to that of Zimbabwe and South Africa.

Runs of homozygosity (ROH) were then explored within medium-high depth genomes, where the fraction of ROH with length >1 Mb (F_{ROH}) affected 3-27% of the genome across all individuals (Fig. 3c). There was no systematic difference in F_{ROH} between red river hogs and bushpigs. Ugandan individuals generally had low levels of F_{ROH} except for one individual, while Ethiopian individuals had relatively long F_{ROH}, supporting the low genetic diversity described within this population. The Madagascar individual had the largest F_{ROH} out of all individuals tested, and exhibited the largest proportion for each length class <10 Mb for all samples excluding Ghana, and <5 Mb for all samples. This is consistent with results comparing linkage disequilibrium (LD) decay between different populations with at least five unrelated individuals, where Madagascar exhibited increased LD (Fig. 3d).

Taken together, these results suggest that the evolutionary histories of red river hogs and bushpigs vary markedly. In light of previous results, we find further evidence that Uganda is likely a region of strong introgression, and that the Ethiopian population underwent strong drift after gene flow with red river hogs. Finally, we find that Malagasy individuals had similar population histories and a level of genetic diversity comparable with other southern bushpig populations, but had increased F_{ROH} and LD.

Bushpig arrival in Madagascar coincides with the expansion of Bantu speakers

The timing of arrival and geographical origin of bushpigs in Madagascar is still unresolved, as previous lines of evidence, e.g. estimated split times and fossil records, appear to be contradictory. We therefore explored the putative founding of this population. We first measured the amount of shared history between the Malagasy population and each of the other populations (Fig. 4a) [43]. Our results suggested that amongst sampled populations, those from South Africa and Zimbabwe have the longest shared history with Madagascar. This was consistent with our results exploring gene flow and connectivity, which showed a weaker

barrier between Madagascar and these two southern populations when compared with other bushpig populations (Fig. 2b,c), the neighbour-joining tree (Fig. 1c) and D_{xy} values (Fig. 2a).

Split times between populations were next examined, including the species split between red river hogs and African bushpigs and the split between bushpig populations on mainland Africa and Madagascar (Fig. 4b) [44]. The species split was estimated to have occurred ~300 kya, and consistent with the outgroup f3 results, Madagascar exhibited the lowest split times with populations in South Africa and Zimbabwe (~850 and ~500 years ago, respectively). This further suggested that either one of these or an unsampled population within the same geographic region was the population of origin. Additionally, we investigated putative recent demographic events for the Madagascar population (Fig. 4c) [45]. This analysis suggested that the Malagasy population experienced a severe bottleneck, likely a result of a founder event between 1-5 kya. This result was also consistent with the high F_{ROH} (Fig. 3c) and the high LD (Fig. 3d) characterising this population.

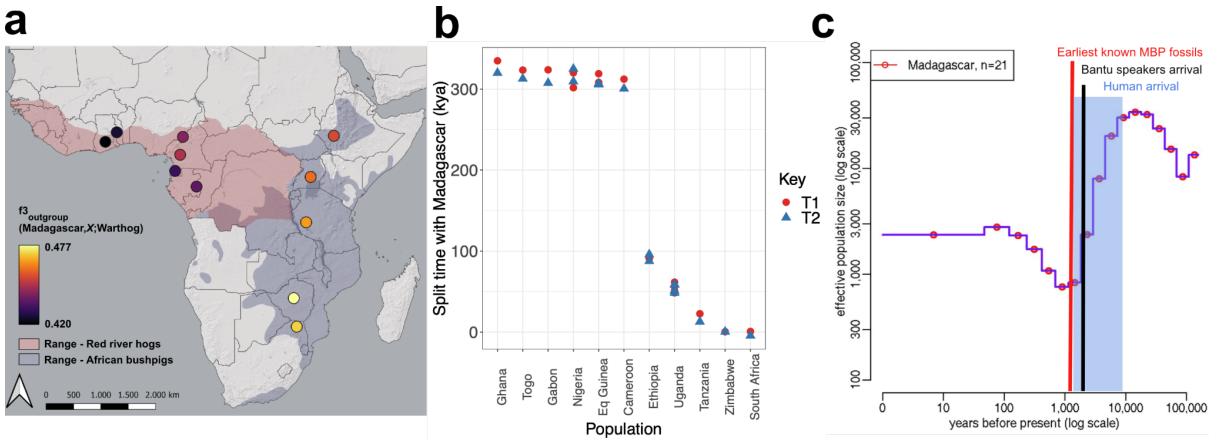

Figure 4. Origin and timing of bushpigs in Madagascar. a) Outgroup f3 statistics in the form f3 (Madagascar, *X*; Warthog), where *X* describes different sampling localities. **b)** Population split times with Madagascar as estimated by the TT method using individual pairs of medium-high depth samples (n = 18). T1 and T2 values, describing population split times, are shown as red and blue points, respectively. A mutation rate of $\mu = 1.49e^{-8}$ per site per generation and a generation time of six years were assumed [4,42] . kya - thousands of years ago. **c)** Recent effective population sizes inferred based on unrelated Madagascar individuals using popSizeABC [45] (n = 21). Shaded region - estimated timing of human arrival in Madagascar [24,26,27]; red line - estimated timing of earliest known bushpig fossils in Madagascar (MBP) [46]; black line - estimated timing of Bantu speakers arrival in Madagascar, as estimated in Pierron *et al.* [26].

Discussion

Biodiversity patterns on the African continent show striking similarities across multiple species, including a division between lineages in W/C Africa and in E/S Africa, with known hybridisation zones spanning across Uganda, South Sudan, and Ethiopia [10,12,13,47]. Although hybridisation between the red river hog and the bushpig has been suggested before, this has not yet been studied in detail and the level of evolutionary divergence between them remains contentious. Moreover, bushpigs also represent the anomaly of being the only large wild mammal which occurs both on mainland Africa and Madagascar, but due to limited data from Malagasy bushpigs, the scenario of colonisation is still largely unknown. Our study is the first to investigate the evolutionary histories of red river hogs and bushpigs at the genome scale, allowing for a better understanding of the processes leading to the formation of two distinct taxa and the colonisation of Madagascar.

Divergence and introgression between red river hogs and bushpigs

Under relatively simple models, we found that the estimated split times between red river hogs and bushpigs could be estimated at ~300 kya (Fig. 4b) and that this time was in the same range as W/C-E/S split time estimates for other African species which show a similar phylogeographic pattern, including African buffalo (~273 kya [10]), baboon (~320 kya) [13], giraffe (~280 kya; submitted), warthog (~226 kya) [4], lion (~245 kya) [11], and spotted hyena (~360 kya) [48]. Although red river hogs and bushpigs are widely considered to be distinct species, in the examples mentioned above, the divergent populations are typically considered to belong to the same species, except for baboons, elephants and with an ongoing taxonomic debate about the species status of giraffes [12,49–51]. In a previous study, Gongora *et al.* [3] estimated the divergence time between red river hogs and bushpigs at 2,710 kya, thus lending further support for a species distinction between red river hogs and bushpigs. However, this estimate was obtained with a wide confidence interval of 200-4,800 kya [3]. Our results suggest (and corroborate recent findings) that divergence times between African suid taxa have thus far been overestimated [4]. This includes red river hogs and bushpigs, where our analyses represent a much younger divergence time and more reticulated evolutionary history than previously known.

Furthermore, our results indicate a complex history of population structure (i.e. a metapopulation or two interconnected metapopulations), with possible periods of increased and decreased connectivity between populations. PSMC curves can be interpreted as representing changes in coalescent effective population size, as is usually done, but this interpretation relies on a very strong assumption of total panmixia [52]. If this assumption is

violated, changes in PSMC curves may alternatively reflect changes in gene flow [53,54]. Thus, an alternative explanation for the observed PSMC curves is that there was a major fragmentation period between 500 kya and 100-200 kya, and a second period more recently, possibly between 100 and 30 kya. Such complex histories could lead to overestimates of divergence times.

Our findings, therefore, have implications for the ongoing taxonomic debate about *Potamochoerus* and for the interpretation of genetic data in this group. The current taxonomic definition of *P. porcus* and *P. larvatus* was primarily based on morphology and a lack of 'convincing evidence' that they interbreed when they come into contact [8], a case which has been previously disputed by previous authors who favour a single-species taxonomy [32,55]. Our results suggest that the two taxa could be a case of incomplete speciation. However, we emphasise that in cases such as *Potamochoerus*, different species concepts might arrive at different conclusions about whether speciation has gone to completion, and we note that taxonomic revisions should draw on various types of data and evidence, e.g. morphology and behaviour [56], which were not considered in the present study.

The impact of changing climate and habitat availability on the complex evolutionary history between red river hog and bushpig is showcased by the Ethiopian population, which has received substantial red river hog gene flow and was fixed for a divergent red river hog mtDNA lineage (Supplementary Fig. S7). Ethiopia is further characterised by a low effective population size and is strongly affected by drift, as illustrated by relatively long ROHs and high F_{ST} values. Hence, Ethiopian bushpigs show contrasting evidence of connectivity and isolation, possibly caused by historical fluctuations in the equatorial forest belt across Africa. These fluctuations could have facilitated intermittent contact and hybridisation between red river hogs and bushpigs as the forest expanded, followed by isolation of resident populations as the forests receded. In line with this, the taxonomic status of *Potamochoerus* in Ethiopia remains unresolved [57] and there is anecdotal evidence of African buffalos in Ethiopia that strongly resemble the forest buffalos of distant central and western Africa [58]. Similarly, the observation of *Potamochoerus* admixture in Uganda coincides with the present-day boundary between two of Africa's major mammalian biogeographical regions, the Guinean-Congolian and the Sudanian core regions [59], an area which also constitutes well-known hybridisation zones for several large-mammal taxa, including elephants [60], and subspecies of buffalo [10] and kob (*Kobus kob*) [61]. However, without more samples from adjacent locations, such as east DRC, the epicentre of hybridisation is speculative.

Origin of Malagasy bushpigs

Our results support the hypothesis that bushpigs were introduced by humans from southeastern Africa into Madagascar ~1,000-1,500 years ago [25,26] and possibly as early as 5,000 years ago. A previous estimate for the most recent common ancestor of mainland and Malagasy bushpigs at 480 kya [29] contradicts this; however, this could be partially caused by problematic temporal calibration and by the limited information contained in mtDNA sequences. Although we cannot pinpoint the precise source population from which the Malagasy bushpigs were introduced with great certainty, our results suggest an origin in southern Africa, as corroborated by the Zimbabwe and South African populations being closer than all other populations when using NGSadmix, evolutionary distances, outgroup f3 statistics, divergence times and EEMS. Our estimate of an effective population size of 1,000 individuals during the bottleneck 1,500 years ago is surprisingly high, assuming that the founder event was a single occurrence involving a limited number of individuals carried to Madagascar by ship. However, the estimate is supported by their observed heterozygosity level, which is similar to levels observed in southern Africa, although we cannot know to what extent southern African populations have been subjected to drift since the founding of the Malagasy population. Multiple introductions, spanning over a longer period of time, or even with animals sourced from different mainland populations (including already admixed individuals), may have inflated this estimate, and we also caution that popSizeABC may not be able to accurately reflect changes in population size happening within a few generations, such as those occurring during a bottleneck with very rapid subsequent regrowth and may also be influenced by population structure.

Despite these limitations, our results provide, to our knowledge, the clearest evidence yet of a recent introduction of bushpigs to Madagascar mediated by humans, most likely populations which started to arrive on Madagascar from southern Africa at least 1,500 years ago and possibly much earlier [26]. Bushpigs likely became established on the island as livestock together with zebu, goats and sheep 700-1,200 years ago based on ^{14}C bone analyses [28], coinciding with the extinction of Madagascar's megafauna likely as the result of human activities such as hunting, pastoralism and farming [62]. Lending strong support to this hypothesis, our dating results are in line with the oldest fossils of Malagasy bushpigs (~1,000 years ago) [46]. Blench [22] hypothesised that human migrants reaching Madagascar must have captured bushpigs in southeastern Africa, introduced them to Madagascar, and made an attempt to domesticate them. Etymological problems over the naming of Malagasy bushpigs (i.e. with a term usually used for bovine in South-East Asia) highlight that there are still outstanding questions regarding the cultural perception and uses of bushpigs in early

Malagasy settlers, composed of both Bantu-speaking and Austronesian-speaking people. Furthermore, the alleged morphological variation between Malagasy subpopulations [5], including the suggestion that they are distinct subspecies [8], had led to suggestions of multiple, distinct introduction pulses through the Comoros Islands and the North Mozambique current [63]. However, from the PCA, NGAadmix and IBS tree we did not identify substantial structure within the island, which is consistent with a relatively homogeneous founder population.

Although we have samples across most of the species' range, we acknowledge that there are some gaps in our inferences. A more even spread across the species ranges, and especially more sampling localities from within the putative suture zone, e.g. southern DR Congo, would increase our understanding of the evolutionary dynamics near the suture zone. In addition, more sampling localities along the East African coastline could help to more precisely identify the source populations for the colonisation of Madagascar [5].

Overall, our study sheds new light on the distribution of genomic diversity and the evolutionary histories of two closely related African pig taxa. It provides yet another example of diverged taxa with a suture zone around western Uganda, as has been shown for numerous other taxa and is characteristic for African mammal phylogeography. The recent split times, moderate and large values of F_{ST} as geographic distance increases and ancestral gene flow between the bushpigs and red river hogs suggest that their evolutionary divergence is young and incomplete, a perspective that should be taken into account in future taxonomic assessments and management plans. Furthermore, our data from Malagasy bushpigs suggest that bushpigs indeed colonised the island by hitchhiking along with the accelerating human colonisation of Madagascar occuring around the onset of the Medieval period. These insights provide answers for long-standing questions regarding the distribution of biodiversity in Africa and the mysterious presence of African bushpigs on Madagascar.

Methods

Sample collection and laboratory protocol

Tissue samples used in this study were sourced from several different collections, detailed in Supplementary Data 1. For non-USDA samples, the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, Valencia, CA, USA) was used for DNA extraction following the manufacturer's protocol. RNase was added to all samples to ensure RNA-free genomic DNA. DNA concentrations were then measured using a Qubit 2.0 Fluorometer and Nanodrop before using gel electrophoresis to check the quality of genomic DNA.

Bushpig hide samples contributed by the USDA were salted, acidified and dried after collection in the field. Samples were purchased by the USDA from willing sellers and stored at -20°C until DNA extraction by standard phenol/chloroform procedures. DNA was dissolved in a solution of 10 mM TrisCl, 1 mM EDTA (TE, pH 8.0) and stored at 4 °C. Sample quality and concentrations were measured by ultraviolet spectrophotometry and double-stranded DNA fluorometry (DeNovix Inc., Wilmington, DE USA; QuantiFluoONE, Promega, Madison, WI, USA).

Sequencing and mapping

All samples were sequenced using Illumina paired-end 150-bp reads. This included 53 samples which were sequenced to low depth (~3–6× depth of coverage) on the Illumina NovaSeq platform and 16 samples sequenced to medium-high depth (~14–49×) on the Illumina HiSeq2500, NextSeq500 or NextSeq2000 platforms (Illumina Inc., San Diego, CA, USA). Sequencing data were then assessed using FastQC [64] and MultiQC [65]. Publicly available data from two red river hog samples from Nigeria were also used in this study (Supplementary Data 1) [66].

The sequencing reads were mapped to the chromosome level assemblies for the common warthog (*Phacochoerus africanus*, accession number: GCA_016906955.1) using a development version of the PALEOMIX BAM pipeline [67] (https://github.com/MikkelSchubert/paleomix; branch 'pub/2022/africa').

Reads were processed using AdapterRemoval v2.3.2 [67] to remove adapter contamination and to merge overlapping reads in order to improve read fidelity. Adapter sequences published by Illumina and BGI were used for trimming. Reads were merged using the *--collapse-conservatively* option, which assigns 'N' to any mismatching position in the alignment for which both bases have the same quality. No trimming of Ns or low-quality bases was performed, and only empty reads resulting from primer-dimers were excluded. Trimmed reads were subsequently mapped using BWA-mem v0.7.17-r1188 [68]. Reads were post-processed using samtools v1.11 [69] commands '*sort*' and '*calmd*', and putative PCR duplicates flagged using the '*markdup*' command and PALEOMIX '*rmdup_collpased*', for paired and unmerged reads respectively.

The resulting BAM alignments were filtered to remove unmapped reads, secondary alignments, PCR duplicates, and supplementary alignments, and reads flagged as having

failed QC. We furthermore removed alignments with an inferred insert size <50 bp or >1,000 bp, and reads where less than 50 bp or 50% were mapped to the reference genome. Finally, we removed pairs of reads mapping to different contigs or in an unexpected orientation and reads for which the mate had been removed by any of the above criteria.

Reference genome and site quality filters

We estimated the mappability of the warthog reference genome using GENMAP [70]. Here, we used 100-bp k-mers allowing for two mismatches (*-K* 100 *-E* 2) and the remaining parameters set to default settings. All sites with a mappability score <1 were excluded from downstream analyses. RepeatMasker v4.1.1 [71] was used to identify repeat elements in the warthog genome assembly, utilising '*rmblast*' as the search engine and '*mammal*' as the query species with default settings. Repeat regions identified with RepeatMasker were masked to limit mismapping in these regions. Annotated sex chromosomes and scaffolds that were not assembled into chromosomes were also excluded.

We also removed genomic regions with unusually high heterozygosity to avoid mismapping artefacts driven by multimapping on paralogous and other repetitive regions. We first estimated genotype likelihoods for SNPs using Angsd [72] with the GATK model (*-GL* 2), -minimum mapping quality of 30 (*-minMapQ* 30), a minimum base quality of 30 (*-minQ* 30), a *p*-value of $1e^{-6}$ to call SNPs (*-snp_pval* $1e^{-6}$) and kept only SNPs with minor allele frequency (MAF) > 0.05 (*-minmaf* 0.05). Genotype likelihoods were then used as input for PCAngsd's per site Hardy-Weinberg equilibrium (HWE) test [73], which estimates inbreeding coefficients (*F*), and a likelihood ratio test statistic (LRT) for evidence of deviation from HWE, while controlling for population structure. The PCAngsd MAP test [73] was also used to select the optimal number of principal components in each case. Sites with $F < -0.9$ and LRT > 24 were subsequently removed as they may have been driven by mapping artefacts, and therefore all regions within 10 kb from such sites were also discarded. We ran this analysis separately for red river hogs and bushpigs samples.

Finally, we removed sites with extreme depth. We estimated the global depth (read count across all samples) for each site using Angsd [72] (*-minMapQ* 30 *-minQ* 30 *-doCounts* 1 *-doDepth* 1 *-dumpCounts* 1 *-maxdepth* 4000). This was done separately for each species for all (n = 67), unrelated (n = 54) and medium-high depth samples (n = 18) (Supplementary Data 1). Only autosomal chromosomes were included. From the global depth we calculated the upper 1% and lower 3% percentiles and visually inspected the plots before deciding on a threshold for excluding sites with extreme sequencing depth. Only sites that were within the

548 thresholds for both low- and medium-high depth samples were used in the downstream
549 analyses.

550 **Sample filters**

551 We identified and excluded samples with high sequencing error rates based on the "perfect
552 individual" approach [74]. The rationale behind this approach is that any sample in the dataset
553 should have equal genetic distance to the outgroup and therefore samples with excess/deficit
554 of derived alleles would be interpreted as errors. As the "perfect individual" we used a high-
555 depth individual from Ghana (BPigGha0038; Supplementary Data 1). This sample was
556 processed with Angsd [72] to create a consensus sequence (*-doFasta* 2) taking the most
557 commonly observed base as the consensus (*-doAncError* 2) while setting the base quality to
558 at least 30 (*-minQ* 30). We chose the common warthog as an outgroup and mapped all
559 samples to the consensus using BWA excluding sex chromosomes, the mitogenome, repeats
560 and sites with mappability <1. Individuals with high error rates (>0.001) were removed from
561 downstream analyses (Supplementary Fig. S1).
562
563 We then considered relatedness between samples, where we identified and removed potential
564 relatives and duplicated samples using the methodology described in IBSRELATE [75]. First,
565 we calculated the Site Allele Frequency (SAF) likelihood in Angsd [72] for each individual. We
566 used the genotype likelihood-based approach assuming HWE (*-doSaf* 1). The warthog
567 genome was used as ancestral reference (*-anc*), a minimum mapping quality of 30 (*-minMapQ*
568 30), a minimum base quality of 30 (*-minQ* 30), and the GATK method (*-GL* 2). Then, we
569 inferred the two-dimensional site frequency spectra (2D-SFS) pairwise among all possible
570 combinations of individuals. To limit the computational time we limited the number of sites
571 surveyed to the first 50,000 sites. Based on the 2D-SFS we calculated: R0, R1 and KING-
572 robust kinship [75,76], which can be used to identify close familiar relatives. For the analysis we
573 combined all the data from bushpig and red river hog in this analysis in order to account for
574 potentially interspecies duplicates or mislabeled samples. We identified and removed an
575 individual from each pair of first and second degree relatives.

576 **Imputation**

577 Imputation was performed using BEAGLE3 [77] from genotype likelihoods (GLs) estimated in
578 Angsd [72]. GLs were estimated using the GATK genotype likelihood model (*-GL* 2) and only
579 keeping sites that had a *p*-value less than $1e^{-6}$ (*-SNP_pval* 1e-6) for being variable in addition
580 to only keeping sites that passed initial QC (*-sites*) as well as using a minimum MAF of 0.025
581 (*-minMAF* 0.025). We assumed the major allele was fixed and the minor was unknown when

582 estimating GLs (*-doMajorMinor* 1 *-doMAF* 2). We further filtered imputation results by only
583 keeping sites with an imputation score $R^2 > 0.95$ and which had a maximum of 5% missingness
584 after applying a > 0.95 posterior probability cutoff on genotype calls.

585 **PCA, IBS and population structure**

586 Beagle GL input files were first generated using Angsd [72], keeping only the sites that passed
587 QC, with additional filters of removing tri-allelic sites, and with a minor allele frequency filter of
588 0.05. We used PCAngsd v1.02 [73] to estimate the covariance matrix and identify potentially
589 population structure for all individuals. A pairwise identity-by-state (IBS) matrix was then
590 generated using Angsd, using the sample filters and including the -doIBS 1 flag. A neighbour-
591 joining tree was then estimated using this matrix using the *ape* library in R [78].

592 **NGSadmix & evalAdmix**

593 Admixture proportions for each population were inferred based on GL using NGSadmix [34]. A
594 Beagle file, using the same filters to investigate population structure with PCAngsd was taken
595 and randomly thinned to contain one million sites for computational practicality. We ran
596 NGSadmix with $K = 2$ to $K = 9$ until the model converged, where the top 3 maximum likelihood
597 runs were within 10 log-likelihood units of each other or until a limit of 4000 independent runs
598 was reached without convergence. $K = 9$ did not converge after 4000 independent runs, likely
599 constrained by the number of samples per population. Model-based analyses of population
600 structure make a set of assumptions about the data (e.g., individuals are unrelated, are in
601 HWE, exhibit no LD, and that each ancestral population is represented by multiple unadmixed
602 individuals with no subsequent drift). Therefore, we calculated the correlations of residuals
603 using evalAdmix [35] for each pair of individuals to evaluate model fit, and to test whether the
604 data violated some of these assumptions for K ancestral clusters.

605 **Population differentiation (F_{ST} / D_{xy})**

606 To quantify the extent of genetic differentiation between red river hog and bushpig populations,
607 we used Hudson's estimator for genome-wide F_{ST} [36]. This analysis encompassed two
608 approaches: one utilising called genotypes for the 18 medium-high depth genomes
609 (Supplementary Fig. S5), and another utilising all 54 unrelated genomes and estimating values
610 from population-level 2D-SFS inferred from genotype likelihoods using winsfs
611 (https://github.com/malthesr/winsfs). We also calculated the absolute genome-wide nucleotide
612 divergence (D_{xy}) for all population pairs using the same approach.

Estimation of effective migration surfaces (EEMS)

To investigate effective migration and gene flow connectivity between populations, we used the Estimated Effective Migration Surfaces (EEMS) program [38]. A distance matrix was created from individual-level 2D-SFS estimated from GLs, and was used as input for the program. EEMS was run using 300 demes for three independent runs of 30 million iterations, discarding the first 15 million as burn-in. Convergence was assessed visually and by using the Gelman–Rubin diagnostic in the *coda* R package [79].

D- and *f*-branch statistics

To explore signatures of introgression between populations of red river hogs and African bushpigs, the Dsuite package [80] was utilised on variable sites of medium-high depth individuals as input, with the topology of a neighbour-joining tree based on pairwise Hudson's F_{ST} between individual pairs using the *ape* library in R [36,78] and the common warthog as an outgroup (Supplementary Data 1). The Dtrios function in Dsuite calculates the D-statistics for all possible trio combinations, which are then used for calculating *f*-branch statistics, using the *f*-branch command. A summary of these results within the provided phylogenetic framework is presented as a heatmap (Supplementary Fig. S6).

PSMC

The Pairwise Sequentially Markovian Coalescent (PSMC) algorithm [40,41] was used to infer changes in historical population sizes by including all individuals sequenced at medium-high depth. PSMC was run with default parameters. In addition to the size quality filter, we also excluded sites based on the average depth per individual divided by three as a minimum, and twice the average depth per individual as a maximum. We used a mutation rate of $1.49e^{-8}$ per site per generation and a generation time of six years, as described for warthogs [4,42].

Heterozygosity

Genetic diversity of pig populations were approximated through the estimation of genome-wide heterozygosity. Individual-level heterozygosity was estimated in Angsd [72] using individual-level site frequency spectra, measured as the proportion of heterozygous loci per sample. The GATK genotype likelihood model was utilised in Angsd (*-GL* 2), with minimum quality filters on mapping (*-minMapQ* 30) and base quality (*-minQ* 30), while reducing the amount of reads with excessive mismatches (*-C* 50).

Runs of homozygosity

Runs of homozygosity (ROH) analyses were performed using PLINK v1.9 [81]. PLINK files included only filtered variable sites within medium-high depth samples (n = 18), with an additional depth filter (10 reads minimum) and at least two heterozygous reads to make a heterozygous call. In order to generate more accurate ROH regions, we further excluded SNPs with MAF < 0.05 (*--maf*) and missing genotype calls (*--geno*) < 0.05. For each individual, we then used PLINK with *--homozyg* to scan the ROH regions, with scanning window modifiers (*--homozyg-window-het* 3 *--homozyg-window-missing* 20). SNP sites with > 50% heterozygous genotypes across individuals were also excluded.

LDdecay

Linkage disequilibrium (LD) decay curves were generated for four populations which included at least five samples (Cameroon, Eq Guinea, Madagascar, and Tanzania) to reduce the potential bias among the variable sample sizes among populations [82]. We calculated LD using the *relate* R package [83] for each population using imputed polymorphic sites from chr16. These sites were thinned to 10% of the original data using PLINK 1.9 (*--thin* 0.1) function [81] to minimise computational time. Pairwise LD was calculated using 36,417 SNPs and 5 Mb physical distance, at which the curves plateaued.

Outgroup f3 statistics

To further test gene flow between the Malagasy population and other red river hog and bushpig populations, outgroup f3 statistics were calculated based on genotype calls from medium-high depth individuals using ADMIXTOOLS2 [43]. f2 statistics were first calculated for each population using five million bp blocks. Using the common warthog as an outgroup, outgroup f3 was estimated in the form of f3 (Madagascar, *X*; Warthog), where *X* represents different populations of red river hogs and bushpigs.

TT and split time estimations

Population split times were estimated from unfolded individual 2D-SFS from genotype calls between medium-high depth samples using the Two-Two (TT) method [44,84], polarised against the common warthog, desert warthog (Supplementary Data 1) and the domestic pig (SRA: SAMN28197093). T1 and T2 values were calculated using formulae described in Sjödin *et al.* [44] using a custom R script with a mutation rate of $1.49e^{-8}$ per site per generation and a generation time of six years as in PSMC analyses [4,42].

PopSizeABC

In order to estimate recent population size changes that cannot be captured by PSMC, we used popSizeABC [45] on imputed data, with a focus on a sufficiently large sample of unrelated individuals (n = 21) in the Malagasy bushpig population. PopSizeABC takes VCF files per chromosome as input, and estimates linkage disequilibrium curves and site frequency spectra for tested populations. PopSizeABC population size estimates require multiple simulations of demographic scenarios to compare a posterior distribution of simulation derived parameters to those observed in the real data. For this analysis, 210,000 simulations were performed for 100 2-Mb regions per simulation as per the suggested settings in the popSizeABC publication for the software. A minimum MAF threshold of 0.1 was applied to calculation of the site frequency spectra and 0.2 for calculation of the linkage disequilibrium curves, again in accordance with the suggested parameters in Boitard *et al.* [45]. The same recombination rate, mutation rate and generation time used in PSMC and TT were used in popSizeABC.

Data availability

Data used in this study are described within the Article, Supplementary Data 1 and Supplementary Information. All data are available upon request.

Acknowledgements

We thank Amal Al-Chaer for her invaluable help with DNA extractions. We are also grateful to Peter Arctander, who organised sample collections between the 1980s and 1990s and to David Moyer for contributing Tanzanian bushpig samples, collected between 1995-1997 and subsequently stored in the collection at the University of Copenhagen. We would also like to acknowledge our collaborators from African wildlife management authorities for granting express permission to use samples within this study. We thank Komlan Afiademanyo (Université de Lomé, Togo), Flobert Njikou (Université de Yaoundé I, Cameroon) and Gabriel Ngua (ANDEGE, Equatorial Guinea) for collecting samples of red river hogs in Western and Central Africa between 2007 and 2010. RFB, XL and IM are supported by a Villum Young Investigator grant (19114) awarded to IM. AA, RFB, LL and ZL are funded by the Novo Nordisk Foundation (NNF20OC0061343). ABO is supported by a Carlsberg Foundation Reintegration Fellowship (CF19-0427). MSR and AA are supported by the Independent Research Fund Denmark (grant number: 8021-00360B). GGE and RH are supported by a Danmarks Frie Forskningsfond Sapere Aude research grant (DFF8049-00098B), and RH is further supported by a Carlsberg Young Researcher grant (CF21-0497). GB, JS, LC, and PG are members of the EDB laboratory, which is supported by the Laboratoire d'Excellence (LABEX) CEBA (grant

ANR-10-LABX-25-01) and LABEX TULIP (grant ANR-10-LABX-0041), both managed by the Agence Nationale de la Recherche in France.

Author information

Contributions

RFB: methodology, analysis, writing - original draft; LDB: methodology, analysis, writing - original draft; ABO: methodology, analysis, writing - review & editing; MSR: methodology & analysis; XL: methodology & analysis; GB: sampling, writing - review & editing; JS: writing - review & editing; CGS: analysis, writing - review & editing; SH: analysis; DZ: writing - review & editing; MP: sampling, writing - review & editing; VM: sampling; CM: sampling; MS: methodology, writing - review & editing; JK: analysis; LQ: analysis; GGE: analysis, writing - review & editing; RR: analysis, MH: analysis; LL: analysis; XW: analysis; MPH: sampling, writing - review & editing; TPLS: sampling, writing - review & editing; KH: analysis; MSS: sampling, writing - review & editing; AA: sampling; LC: writing - review & editing; CR: analysis, writing - review & editing; PG: sampling, writing - review & editing; HRS: sampling, writing - review & editing; IM: supervision, writing - original draft; AA: supervision, writing - original draft, RH: supervision, writing - original draft. All authors proofread and approved the final version of the manuscript.

Ethics declarations

Competing interests

The authors declare no competing interests.

Materials and correspondence

Materials requests and correspondence: Renzo F. Balboa (renzo.balboa@bio.ku.dk), Laura D. Bertola (laura.bertola@bio.ku.dk), Ida Moltke (ida@bio.ku.dk), Anders Albrechtsen (aalbrechtsen@bio.ku.dk), and Rasmus Heller (rheller@bio.ku.dk).

References

1. Frantz, L. *et al.* The evolution of Suidae. *Annu. Rev. Anim. Biosci.* **4**, 61–85 (2016).
2. Kingdon, J. *Mammals of Africa: Volume VI: Pigs, Hippopotamuses, Chevrotain, Giraffes, Deer and Bovids*. (Bloomsbury Natural History, 2013).
3. Gongora, J. *et al.* Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). *Zool. Scr.* **40**, 327–335 (2011).
4. Garcia-Erill, G. *et al.* Warthog genomes resolve an evolutionary conundrum and reveal introgression of disease resistance genes. *Mol. Biol. Evol.* **39**, (2022).
5. Oliver, W. L. R., IUCN/SSC Pigs and Peccaries Specialist Group & IUCN/SSC Hippo Specialist Group. *Pigs, Peccaries, and Hippos: Status Survey and Conservation Action Plan*. (IUCN, 1993).
6. Halternorth, T. & Diller, H. *A field guide to the Mammals of Africa including Madagascar*. (W. Collins, 1980).
7. Grubb, P. Controversial scientific names of African mammals. *Afr. Zool.* **39**, 91–109 (2004).
8. Grubb, P. *The Afrotropical suids (Phacochoerus, Hylochoerus and Potamochoerus). In Pigs, peccaries, and hippos: status survey and conservation action plan*. (International Union for Conservation of Nature and Natural Resources, 1993).
9. Kingdon, J. *Mammals of Africa: Volume I: Introductory Chapters and Afrotheria*. (Bloomsbury Natural History, 2013).
10. Smitz, N. *et al.* Pan-African genetic structure in the African buffalo (*Syncerus caffer*): investigating intraspecific divergence. *PLoS One* **8**, e56235 (2013).
11. Bertola, L. D. *et al.* Phylogeographic patterns in Africa and high resolution delineation of genetic clades in the lion (*Panthera leo*). *Sci. Rep.* **6**, 30807 (2016).
12. Rohland, N. *et al.* Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* **8**, e1000564 (2010).
13. Zinner, D., Groeneveld, L. F., Keller, C. & Roos, C. Mitochondrial phylogeography of baboons (*Papio spp.*): indication for introgressive hybridization? *BMC Evol. Biol.* **9**, 83 (2009).
14. Ghiglieri, M. P. *et al.* Bush pig (*Potamochoerus porcus*) polychromatism and ecology in Kibale Forest, Uganda. *Afr. J. Ecol.* **20**, 231–236 (1982).
15. Bonnald, J. *et al.* Who are the elephants living in the hybridization zone? How genetics may guide conservation to better protect endangered elephants. *Glob. Ecol. Conserv.* **32**, e01917 (2021).
16. Garbutt, N. *Mammals of Madagascar*. (Pica Press, 1999).
17. Raval, U. & Veeraswamy, K. India-Madagascar separation: breakup along a pre-existing mobile belt and chipping of the craton. *Gondwana Res.* **6**, 467–485 (2003).

769 18. Rabinowitz, P. D., Coffin, M. F. & Falvey, D. The separation of Madagascar and Africa.
770 *Science* **220**, 67–69 (1983).
771 19. McCall, R. A. Implications of recent geological investigations of the Mozambique
772 Channel for the mammalian colonization of Madagascar. *Proc. Biol. Sci.* **264**, 663–665 (1997).
773 20. Ali, J. R. & Hedges, S. B. The colonisation of Madagascar by land-bound vertebrates.
774 *Biol. Rev. Camb. Philos. Soc.* (2023) doi:10.1111/brv.12966.
775 21. Yoder, A. D., Cartmill, M., Ruvolo, M., Smith, K. & Vilgalys, R. Ancient single origin for
776 Malagasy primates. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5122–5126 (1996).
777 22. Blench, R. New palaeozoogeographical evidence for the settlement of Madagascar.
778 *Azania* **42**, 69–82 (2007).
779 23. Boivin, N., Crowther, A., Helm, R. & Fuller, D. Q. East Africa and Madagascar in the
780 Indian Ocean world. *J. World Prehist.* **26**, 213–281 (2013).
781 24. Douglass, K. *et al.* A critical review of radiocarbon dates clarifies the human settlement
782 of Madagascar. *Quat. Sci. Rev.* **221**, 105878 (2019).
783 25. Anderson, A. *et al.* New evidence of megafaunal bone damage indicates late
784 colonization of Madagascar. *PLoS One* **13**, e0204368 (2018).
785 26. Pierron, D. *et al.* Genomic landscape of human diversity across Madagascar. *Proc.*
786 *Natl. Acad. Sci. U. S. A.* **114**, E6498–E6506 (2017).
787 27. Hansford, J. *et al.* Early Holocene human presence in Madagascar evidenced by
788 exploitation of avian megafauna. *Sci Adv* **4**, eaat6925 (2018).
789 28. Hixon, S. W. *et al.* Late Holocene spread of pastoralism coincides with endemic
790 megafaunal extinction on Madagascar. *Proc. Biol. Sci.* **288**, 20211204 (2021).
791 29. Lee, C. *et al.* Genetic origins and diversity of bushpigs from Madagascar
792 (*Potamochoerus larvatus*, family Suidae). *Sci. Rep.* **10**, 20629 (2020).
793 30. Forsyth, C. I. 5. On the species of *Potamochoerus*, the bush-pigs of the Ethiopian
794 region. *Proc. Zool. Soc. Lond.* **65**, 359–370 (1897).
795 31. Rakotoarivony, R. *et al.* Bushpig (*Potamochoerus larvatus*) hunting in rural areas of
796 Madagascar and its health and socioeconomic implications. *Front. Conserv. Sci.* **3**, (2022).
797 32. Seydack, A. H. W. *Potamochoerus larvatus*. *The IUCN Red List of Threatened Species*
798 https://www.iucnredlist.org/species/41770/44140926 (2016).
799 33. Reyna, R., Jori, F., Querouil, S. & Leus, K. *Potamochoerus porcus*. *The IUCN Red List*
800 *of Threatened Species* https://www.iucnredlist.org/species/41771/100469961 (2016).
801 34. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture
802 proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
803 35. Garcia-Erill, G. & Albrechtsen, A. Evaluation of model fit of inferred admixture
804 proportions. *Mol. Ecol. Resour.* **20**, 936–949 (2020).
805 36. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from

806 DNA sequence data. *Genetics* **132**, 583–589 (1992).
807 37. Nei, M. & Miller, J. C. A simple method for estimating average number of nucleotide
808 substitutions within and between populations from restriction data. *Genetics* **125**, 873–879
809 (1990).
810 38. Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with
811 estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016).
812 39. Zheng, Y. & Janke, A. Gene flow analysis method, the D-statistic, is robust in a wide
813 parameter space. *BMC Bioinform.* **19**, 10 (2018).
814 40. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from
815 genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
816 41. Li, H. & Durbin, R. Inference of human population history from individual whole-
817 genome sequences. *Nature* **475**, 493–496 (2011).
818 42. Pacifici, M. *et al.* Generation length for mammals. *Nat. Conserv.* **5**, 89–94 (2013).
819 43. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093
820 (2012).
821 44. Sjödin, P., McKenna, J. & Jakobsson, M. Estimating divergence times from DNA
822 sequences. *Genetics* **217**, (2021).
823 45. Boitard, S., Rodríguez, W., Jay, F., Mona, S. & Austerlitz, F. Inferring population size
824 history from large samples of genome-wide molecular data - an approximate Bayesian
825 computation approach. *PLoS Genet.* **12**, e1005877 (2016).
826 46. Rakotozafy, L. M. A. & Goodman, S. M. Contribution à l'étude zooarchéologique de la
827 région du Sud-ouest et extrême Sud de Madagascar sur la base des collections de l'ICMAA
828 de l'Université d'Antananarivo. *Taloha* 14–15 (2005).
829 47. Bertola, L. D. *et al.* Whole genome sequencing and the application of a SNP panel
830 reveal primary evolutionary lineages and genomic variation in the lion (*Panthera leo*). *BMC*
831 *Genom.* **23**, 321 (2022).
832 48. Rohland, N. *et al.* The population history of extant and extinct hyenas. *Mol. Biol. Evol.*
833 **22**, 2435–2443 (2005).
834 49. Bercovitch, F. B. *et al.* How many species of giraffe are there? *Curr. Biol.* vol. 27 R136–
835 R137 (2017).
836 50. Coimbra, R. T. F. *et al.* Whole-genome analysis of giraffe supports four distinct
837 species. *Curr. Biol.* **31**, 2929–2938.e5 (2021).
838 51. Petzold, A. & Hassanin, A. A comparative approach for species delimitation based on
839 multiple methods of multi-locus DNA sequence analysis: a case study of the genus *Giraffa*
840 (Mammalia, Cetartiodactyla). *PLoS One* **15**, e0217956 (2020).
841 52. Mazet, O., Rodríguez, W., Grusea, S., Boitard, S. & Chikhi, L. On the importance of
842 being structured: instantaneous coalescence rates and human evolution—lessons for

ancestral population size inference? *Heredity* **116**, 362–371 (2015).

53. Chikhi, L. *et al.* The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity* **120**, 13–24 (2018).

54. Arredondo, A. *et al.* Inferring number of populations and changes in connectivity under the n-island model. *Heredity* **126**, 896–912 (2021).

55. Leslie, D. M. & Huffman, B. A. *Potamochoerus porcus* (Artiodactyla: Suidae). *Mamm. Species* **47**, 15–31 (2015).

56. Ravinet, M. *et al.* Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* **30**, 1450–1477 (2017).

57. Melletti, M. *et al.* Red River hog *Potamochoerus porcus* (Linnaeus, 1758). in *Ecology, Conservation and Management of Wild Pigs and Peccaries* 134–149 (Cambridge University Press, 2017).

58. Roussos, N. N. Red forest buffalo in Kaffa Province. *Walia* **1975**, 16–17 (1975).

59. Linder, H. P. *et al.* The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. *J. Biogeogr.* **39**, 1189–1205 (2012).

60. Mondol, S. *et al.* New evidence for hybrid zones of forest and savanna elephants in Central and West Africa. *Mol. Ecol.* **24**, 6134–6147 (2015).

61. Lorenzen, E. D., De Neergaard, R., Arctander, P. & Siegismund, H. R. Phylogeography, hybridization and Pleistocene refugia of the kob antelope (*Kobus kob*). *Mol. Ecol.* **16**, 3241–3252 (2007).

62. Godfrey, L. R. *et al.* A new interpretation of Madagascar's megafaunal decline: The 'Subsistence Shift Hypothesis'. *J. Hum. Evol.* **130**, 126–140 (2019).

63. Ray, H. P. & Alpers, E. A. *Cross currents and community networks: the history of the Indian Ocean world*. (Oxford University Press New Delhi, 2007).

64. Andrews, S. *FastQC: A quality control tool for high throughput sequence data*. (2010).

65. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047 (2016).

66. Xie, H.-B. *et al.* African suid genomes provide insights into the local adaptation to diverse African environments. *Mol. Biol. Evol.* **39**, (2022).

67. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).

68. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

69. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, (2021).

70. Pockrandt, C., Alzamel, M., Iliopoulos, C. S. & Reinert, K. GenMap: ultra-fast

880 computation of genome mappability. *Bioinformatics* **36**, 3687–3692 (2020).
881 71. Smit, A. F. A., Hubley, R. & Green, P. Repeat-Masker Open-3.0.
882 *http://www.repeatmasker.org* (2004).
883 72. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
884 Generation Sequencing Data. *BMC Bioinform.* **15**, 356 (2014).
885 73. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions
886 in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
887 74. Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an
888 early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
889 75. Waples, R. K., Albrechtsen, A. & Moltke, I. Allele frequency-free inference of close
890 familial relationships from genotypes or low-depth sequencing data. *Mol. Ecol.* **28**, 35–48
891 (2019).
892 76. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association
893 studies. *Bioinformatics* **26**, 2867–2873 (2010).
894 77. Ayres, D. L. *et al.* BEAGLE 3: improved performance, scaling, and usability for a high-
895 performance computing library for statistical phylogenetics. *Syst. Biol.* **68**, 1052–1061 (2019).
896 78. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution
897 in R language. *Bioinformatics* **20**, 289–290 (2004).
898 79. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and
899 output analysis for MCMC. *R News* **6**, 7–11 (2006).
900 80. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite - Fast D-statistics and related
901 admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
902 81. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
903 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
904 82. Weir, B. S. & Hill, W. G. Effect of mating structure on variation in linkage disequilibrium.
905 *Genetics* **95**, 477–488 (1980).
906 83. Albrechtsen, A. *et al.* Relatedness mapping and tracts of relatedness for genome-wide
907 data in the presence of linkage disequilibrium. *Genet. Epidemiol.* **33**, 266–274 (2009).
908 84. Schlebusch, C. M. *et al.* Southern African ancient genomes estimate modern human
909 divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).