Warthog Genomes Resolve an Evolutionary Conundrum and Reveal Introgression of Disease Resistance Genes

Genís Garcia-Erill [iD],[†,1] Christian H.F. Jørgensen,[†,1] Vincent B. Muwanika,[2] Xi Wang,[1] Malthe S. Rasmussen,[1] Yvonne A. de Jong,[3] Philippe Gaubert,[4] Ayodeji Olayemi,[5] Jordi Salmona,[4] Thomas M. Butynski,[3] Laura D. Bertola,[1] Hans R. Siegismund,[1] Anders Albrechtsen,[1] and Rasmus Heller*,[1]

[1]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark

[2]Department of Environmental Management, Makerere University, PO Box 7062, Kampala, Uganda

[3]Eastern Africa Primate Diversity and Conservation Program & Lolldaiga Hills Research Programme, PO Box 149, Nanyuki 10400, Kenya

[4]Laboratoire Évolution & Diversité Biologique, Université Toulouse III Paul Sabatier, 31062 Toulouse, France

[5]Natural History Museum, Obafemi Awolowo University, HO 220005 Ile Ife, Nigeria

*Corresponding author: E-mail: rheller@bio.ku.dk.

[†]These authors contributed equally to this work.

Associate editor: Dr Maria C. Ávila-Arcos

Abstract

African wild pigs have a contentious evolutionary and biogeographic history. Until recently, desert warthog (*Phacochoerus aethiopicus*) and common warthog (*P. africanus*) were considered a single species. Molecular evidence surprisingly suggested they diverged at least 4.4 million years ago, and possibly outside of Africa. We sequenced the first whole-genomes of four desert warthogs and 35 common warthogs from throughout their range. We show that these two species diverged much later than previously estimated, 400,000–1,700,000 years ago depending on assumptions of gene flow. This brings it into agreement with the paleontological record. We found that the common warthog originated in western Africa and subsequently colonized eastern and southern Africa. During this range expansion, the common warthog interbred with the desert warthog, presumably in eastern Africa, underlining this region's importance in African biogeography. We found that immune system–related genes may have adaptively introgressed into common warthogs, indicating that resistance to novel diseases was one of the most potent drivers of evolution as common warthogs expanded their range. Hence, we solve some of the key controversies surrounding warthog evolution and reveal a complex evolutionary history involving range expansion, introgression, and adaptation to new diseases.

Key words: *Phacochoerus* evolution, introgression, disease resistance, African phylogeography, population structure.

Introduction

The genus *Phacochoerus* has a contentious and complex taxonomic history. Two species of warthog, the desert warthog (*P. aethiopicus*) and the common warthog (*P. africanus*), were recognized since 1788. Subsequent zoologists, however, erroneously assumed that all extant warthogs belonged to *P. africanus* until Grubb (1993) established that the extant Somali warthog (*P. aethiopicus delamerei*) is conspecific with the extinct Cape warthog (*P. aethiopicus aethiopicus*), meaning there are in fact two extant species of warthog: the common and the desert warthog. Two studies added to this evolutionary conundrum by estimating a surprisingly ancient divergence time between the two species of warthogs: 4.4 Ma (Randi et al. 2002) and 8.8–5.7 Ma (Randi et al. 2002; Gongora et al. 2011). These dates are inconsistent with the fossil evidence, which indicates that *Phacochoerus* first appeared either around 1.0–0.8 Ma (White and Harris 1977; Harris and White 1979;

Harris and Cerling 2002; Souron 2017), or around 2.2–2.0 Ma (Hopwood and Hollyfield 1954; Ewer 1956; Cooke 1994; Pickford 2006, 2012, 2013a, b; Pickford and Gommery 2016). This inconsistency between the genetic data and the fossil record has shrouded the evolutionary relationship between the two species of warthog in controversy, showcasing a more general conflict between evolutionary chronology based on genetic data and fossils (Yang and Donoghue 2016). Moreover, the unresolved divergence time has raised fundamental biogeographic questions, as the earliest known Suinae fossil from Africa is dated to 5.1 Ma (Pickford and Gommery 2016, 2020). This means that the existing genetic estimates for the appearance of the two extant species of *Phacochoerus* overlap with or predate the earliest known occurrence of Suinae in Africa. This, controversially, suggests that the two extant warthogs evolved outside Africa (Gongora et al. 2011).

In addition to the controversial evolutionary relationship between the two warthog species, they remain

Open Access

Article

understudied from a population genetic perspective. Muwanika et al. (2003) used microsatellites and mtDNA to infer three common warthog refugia (western, eastern, and southern Africa). This is consistent with the prevailing phylogeographic pattern for large African mammals (Hewitt 2004; Lorenzen et al. 2012). Lorenzen et al. (2012) built on these results and suggested that eastern Africa was colonized from a southern African refugium, a pattern emerging in many other savanna ungulates including impala (*Aepyceros melampus*; Nersting and Arctander 2001; Lorenzen et al. 2006), wildebeest (*Connochaetes* spp.; Arctander et al. 1999), greater kudu (*Tragelaphus strepsiceros*; Nersting and Arctander 2001), and plains zebra (*Equus quagga*; Pedersen et al. 2018).

Warthogs, in contrast to all other suids, are highly adapted to the open grasslands of the African savannas (Cumming 2013; Grubb and D'Huart 2013; Butynski and De Jong 2018; De Jong and Butynski 2018). Their open-country adaptations include longer legs for cursorial locomotion, a highly specialized dentition, and a large head with broad vision. The adaptation to African savanna habitat from an ancestral forest-adapted stock mirrors those of early hominins relative to their great ape ancestors, and the possible correlation between African suid and hominid evolution and biogeography has been pointed out (White and Harris 1977). Hence, an increased understanding of warthog phylogeography is desirable to enhance our understanding of the biogeographic theater in which humans evolved.

Of the two extant species of warthog, the common warthog occupies a wide range of habitats and has by far the greatest present distribution (fig. 1A). Four subspecies of common warthog are currently recognized (*africanus*, *aeliani*, *massaicus*, and *sundevallii*; Grubb 1993, 2005), but the limits of their geographic distributions are poorly understood and their taxonomic arrangement is in need of validation (Butynski and De Jong 2018). Desert warthogs are currently restricted to Ethiopia, Kenya, and Somalia where they are sympatric with common warthogs in several regions (fig. 1A; De Jong and Butynski 2018, 2021; Butynski and De Jong 2021; De Jong et al. in press). Desert warthogs are adapted to low-lying, arid habitats, whereas common warthogs are typically associated with more moist savanna habitats and open woodlands (De Jong and Butynski 2018). The literature is conflicted as to whether hybridization between common and desert warthogs occurs. Souron (2016) found atypical warthog skulls in the Horn of Africa that might indicate hybrids, but De Jong and Butynski (2018) argued that their ancient divergence time makes hybridization unlikely. These unresolved questions and conflicting observations currently prevent any resolution of the evolutionary history of *Phacochoerus*, as well as pose outstanding questions regarding African suid biogeography and conservation.

In this study, we present the first whole-genome data from warthogs and use it to resolve some of these outstanding questions regarding warthog evolution. To do so, we investigate the genetic structure and

phylogeography within *Phacochoerus*. We particularly focus on estimating the divergence time between the two extant warthog species and identifying possible introgression between them.

Results

After sample filtering, we analyzed 35 common warthogs and four desert warthogs (supplementary material table S1, Supplementary Material online, fig. 1A). Six samples, including one desert warthog, were sequenced to high-depth (∼17×) and the remaining samples were sequenced to low-depth (∼3×). An mtDNA tree with two deeply divergent clades confirmed the taxonomic status of our samples as desert and common warthogs (supplementary material fig. S1, Supplementary Material online). After strict filtering of genomic sites, excluding repeats, regions with low mappability, sites with outlying depth patterns, and regions inferred to be exclusively heterozygous, we retained 1.3 Gbp of the autosomal genome for further analysis.

Population Structure and Phylogeography of the Common Warthog

From the filtered autosomal sites, we visualized the structure within common warthog populations using PCAngsd. The observed genetic structure clustered the samples in discrete groups according to the country from which the samples were obtained (fig. 1B). Ghana was separated from all other populations in PC1, suggesting a deep split between Ghana and the other localities. The structure was also recovered in NGSadmix, where each inferred admixture component corresponds to a country at $K = 5$, with the exception of the single Zambian sample (fig. 1C) which appears as genetically intermediate between Tanzania and Zimbabwe. This is expected from biogeography and is in line with the Principal Component Analysis (PCA). In the Kenyan population there is some genetic substructure, with one sample being modeled as a mixture of the other Kenyan and the Tanzanian clusters, in accordance with the two sampling localities (fig. 1A; supplementary material table S1, Supplementary Material online). The mixed ancestry in one of the Kenyan samples and the single Zambian sample are likely the result of having too few samples from these localities, rather than true admixture between the populations. Evaluation of the admixture proportions using evalAdmix corroborated that five clusters are needed to accurately model population structure (supplementary material figs. S2 and S3, Supplementary Material online). It also indicates that there is some substructure or cryptic relatedness within most of the inferred populations, especially among the Zimbabwean samples. The admixture analyses did not converge for values of $K > 5$. This might be due to sample sizes that are too small to characterize more subtle substructure.

Based on the inferred population structure, we used the countries as population groupings for Treemix. We found the Ghanian population to be the most basal split within the common warthogs, with a progressive splitting of the

Fig. 1. Sampling localities and population structure. (A) Sampling localities and number of individuals remaining after filtering (see supplementary material table S1, Supplementary Material online). Approximate geographic limits of the four currently recognized subspecies of common warthog are shown. Species and subspecies ranges based on Vercammen and Mason (1993), Muwanika et al. (2003), Butynski and De Jong (2018), De Jong and Butynski (2018), De Jong et al. (2018, in press). (B) Plot of common warthog samples, colored by sample country, on the first two principal components inferred with PCAngsd. (C) Admixture proportions of common warthog samples, estimated with NGSadmix assuming five ancestral clusters. (D) Genome-wide heterozygosity of all desert and common warthog samples, calculated from genotype likelihoods with realSFS. Similar levels were obtained for the high-depth individuals with genotype calls (supplementary material fig. S4, Supplementary Material online). Above the plot, we show the topology of the TreeMix tree without migrations (see supplementary material fig. S8, Supplementary Material online for full TreeMix result).
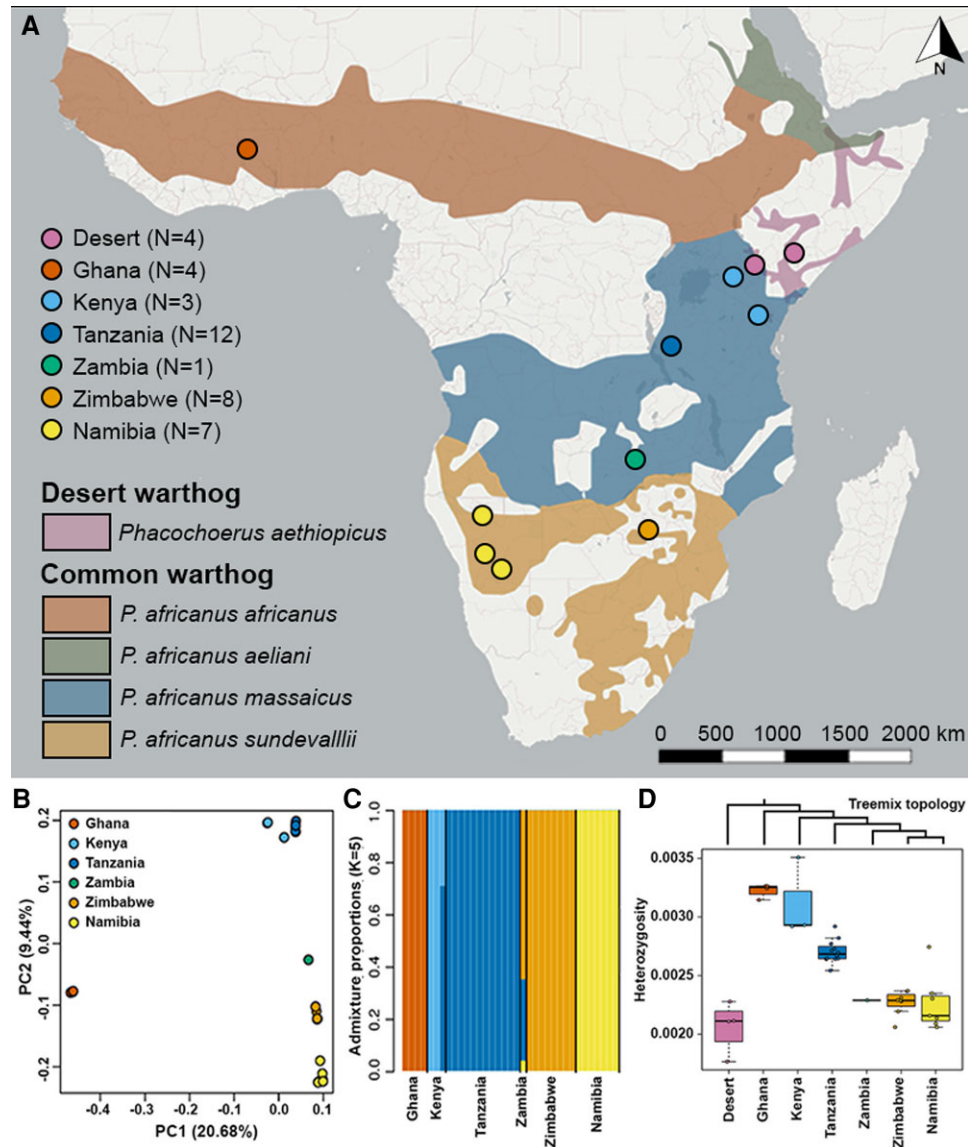
remaining populations from east to south. This progressive splitting is accompanied by a decrease in heterozygosity along an axis from west to east and southwards (fig. 1D, supplementary material fig. S4, Supplementary Material online). This suggests serial founder events as the common warthog colonized new areas. Two outlying samples, one each in Kenya and Namibia, diverged from this pattern by having higher heterozygosity than expected given the position of their population in the expansion. We correlated sample heterozygosity with error rates. This revealed that the high heterozygosity of these two samples is an artifact of their higher per base sequencing error rates (supplementary material fig. S5, Supplementary Material online).

We summarize the phylogeographic interpretation of our analyses in figure 2A. We inferred a range expansion in common warthogs from an origin in western Africa, colonizing first eastern Africa and subsequently southern Africa. The range expansion followed a semi-circular path circumnavigating the Central African rainforest, as shown by the geographically aware population structure

analysis which identifies the rainforest as a strong barrier to gene flow (fig. 3B). This stepwise range expansion was corroborated by a directionality test (supplementary material table S2, Supplementary Material online). We quantified the genetic distance between populations using F_{ST} and obtained similar results when using single high-depth–sequenced individuals as when using genotype likelihoods (GLs) for groups of low-depth–sequenced individuals (fig. 2C). The inferred F_{ST} between common warthog populations was as low as 0.06 between Kenya and Tanzania and up to 0.34 between Namibia and Ghana. Overall, genetic differentiation between Ghana and all other common warthog populations ("eastern and southern African" or "ESA" in the following) was high, in agreement with the PCA and mtDNA tree (supplementary material fig. S1, Supplementary Material online). Despite nominally belonging to different subspecies (fig. 1A), Tanzania and Zimbabwe had a low F_{ST} of 0.08. This is almost on par with that between Zimbabwe and Namibia ($F_{ST} = 0.07$), which nominally belong to the
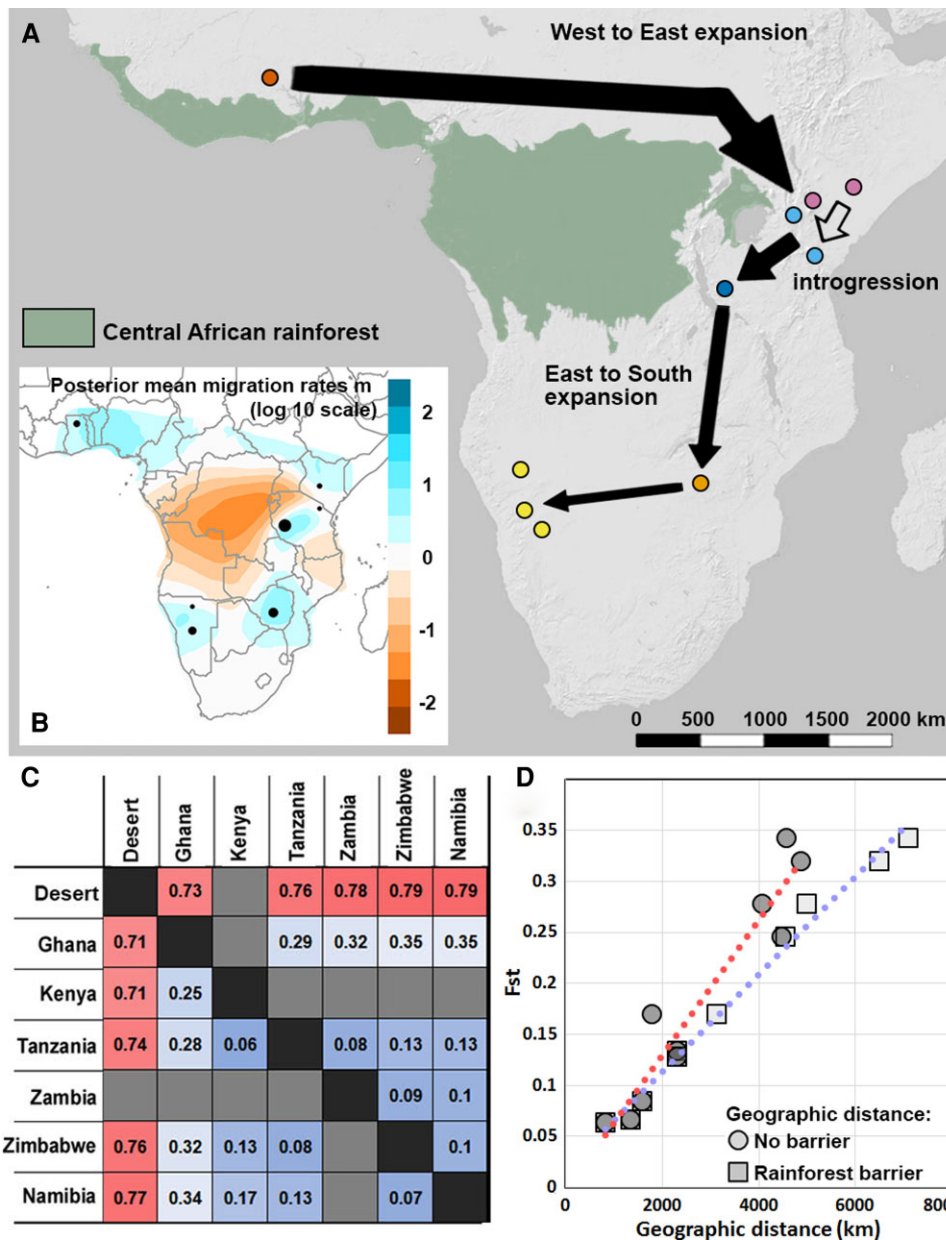
FIG. 2. Warthog phylogeographic synthesis. (A) Summary of the main phylogeographic findings in the present study. Solid arrows show the inferred directionality of the range expansions; their width is proportional to inferred genetic diversity. The transparent arrow marks the introgression between desert warthogs and common warthogs. Also shown is an approximate current outline of the Central African rainforest (FAO 2012). (B) Posterior mean migration rates among common warthog localities, estimated with EEMS. (C) Population pairwise F_{ST} values based on single, high-depth individuals (above diagonal), and several low-depth individuals (below diagonal) per population. (D) F_{ST} between pairs of common warthog populations, estimated with realSFS, against the geographical distance between corresponding pairs of localities. Geographic distance is the shortest distance when either taking into account the Central African rainforest as a barrier (squares) or taking the great circle distance among localities (circles). Notice the improved linear fit when including the rainforest as a barrier, compared with the great circle distance.

same subspecies (fig. 1A). Finally, the range expansion is further supported by a linear relationship between pairwise F_{ST} and geographic distance when following a trajectory around the Central African forest (fig. 3D). Common warthog phylogeography is therefore closely tied to the distribution of savanna habitat, and it is probable that extensive Central African forest cover prevented a southward expansion of common warthogs for prolonged periods of the Pleistocene.

Desert warthog samples in general showed lower genetic diversity than common warthogs, although their heterozygosity values overlap with common warthogs from southern African populations (fig. 1D). F_{ST} was 0.71–0.77 between desert and common warthog populations (fig. 2C). When high, these F_{ST} values indicate that both species still share a substantial amount of genetic variation despite their presumed ancient divergence time.

Introgression between Species

A population structure analysis combining common and desert warthog samples did not identify recent hybrids between the two species (supplementary material fig. S6, Supplementary Material online). We investigated historical gene flow by using D-statistics and found a strong signal of gene flow between desert warthog and all ESA common warthog populations relative to Ghana (fig. 3A). We verified that desert warthog gene flow was not an artifact of gene flow from an out-group into Ghana (supplementary material table S3, Supplementary Material online). The similar magnitude of introgression in all ESA common warthog populations suggests that most of this admixture occurred in a population ancestral to the ESA common warthog populations, that is, at an early stage of their range expansion. One low-depth sample (sample ID 8931) from Tsavo West National Park, an
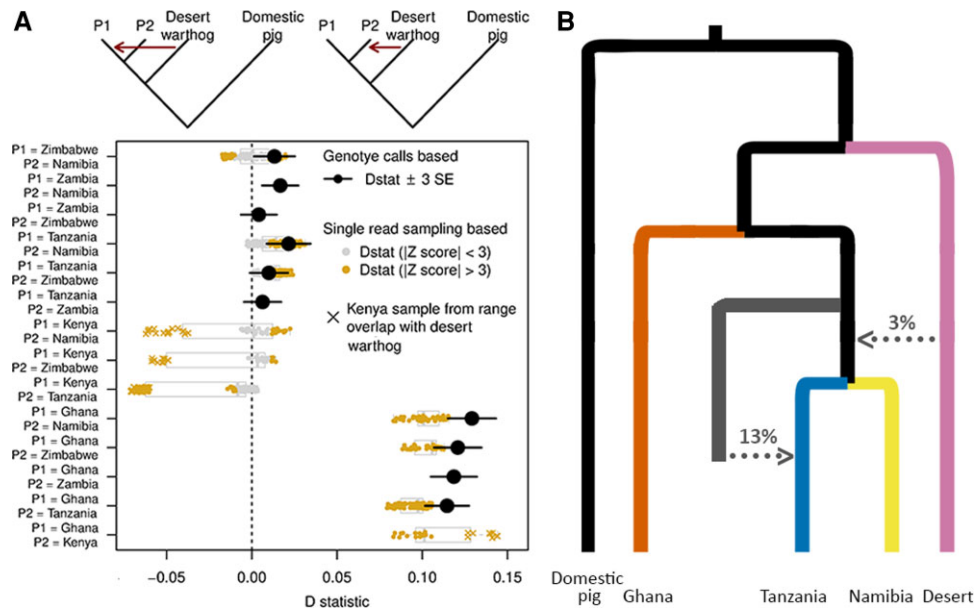
FIG. 3. Introgression between desert warthogs and common warthogs. (*A*) D-statistics when using desert warthog samples as P3 and common warthog samples as P1 and P2, grouped by the country of origin of the two common warthog samples. D-statistics are estimated both from called genotypes for the high-depth samples, using qpDstat, and by single read sampling for all possible combinations of low-depth samples from the corresponding P1, P2, and P3 populations, using ANGSD. (*B*) Cartoon visualization of an admixture graph fitted to a reduced data set, showing the relationships among common warthogs from the three main lineages and the desert warthog. The admixture graph was estimated with qpGraph from called genotypes using a single high-depth sample per population. See in supplementary material fig. S7, Supplementary Material online estimates of branch length for the graph, and all other compatible admixture graphs for the same set of populations identified with qpBrute.

area of sympatry in Kenya, shows excessive allele sharing with desert warthogs compared with other ESA common warthog samples (fig. 3A). This might suggest gene flow that is too old to be detectable in the admixture analysis, but recent enough to occur after the divergence between Kenyan populations.

Using qpGraph, we identified 28 admixture graphs without significant outliers ($|Z| < 3$). When using a topology consistent with the inferred phylogeography, where Ghana is an out-group to all other common warthog populations, all admixture graphs show gene flow between common warthogs and desert warthogs (supplementary material fig. S7, Supplementary Material online). We chose the best-fitting graph as the one with the lowest maximum absolute Z-score of the difference between observed and fitted f_4 statistic. This graph includes 3% introgression from a desert warthog population into a population ancestral to Tanzania and Namibia, and 13% introgression from a ghost common warthog population into Tanzania (fig. 3B). Similarly, when migrations were modeled in TreeMix, we found migration edges between desert and common warthogs at $m = 2$ and $m = 3$, albeit with inconsistent placement of the migration edges (supplementary material fig. S8, Supplementary Material online).

Demographic History

We estimated historical effective population sizes for the high-depth individuals using the Pairwise Sequentially Markovian Coalescent (PSMC). The three Ghanaian samples showed a different demographic history than the ESA common warthog populations (fig. 4A), even in the distant past. We attribute this to admixture between ESA common warthogs and desert warthogs. The desert warthog showed slightly different effective population sizes compared with common warthog populations around 900–200 thousand years ago (kya), but after 200 kya, the desert warthog and ESA common warthogs had very similar population sizes. In contrast, from 200 kya onwards, the PSMC strongly suggests that Ghana had an increase in effective population size, whereas all other populations underwent population declines.

To better understand the time frame of the major evolutionary events in *Phacochoerus*, we used fastsimcoal2 to estimate divergence times and other demographic parameters based on two-dimensional site frequency spectra (2dSFS). We initially evaluated a set of three models with different assumptions on the admixture between common and desert warthogs, based on the qpGraph results (supplementary material fig. S9, Supplementary Material online). Out of these three models, the "1-admixture" model (fig. 4B), which corresponds to the best-fitting qpGraph, had the best fit (supplementary material table S4, Supplementary Material online). The time of the basal split between desert and common warthogs was estimated under this model to be 473 kya (95% confidence interval [CI] 390–624 kya; fig. 4B, supplementary material table S5, Supplementary Material online). However, qpGraph cannot model bidirectional gene flow. Since some accepted admixture graphs included gene flow from common warthogs to desert warthogs (supplementary
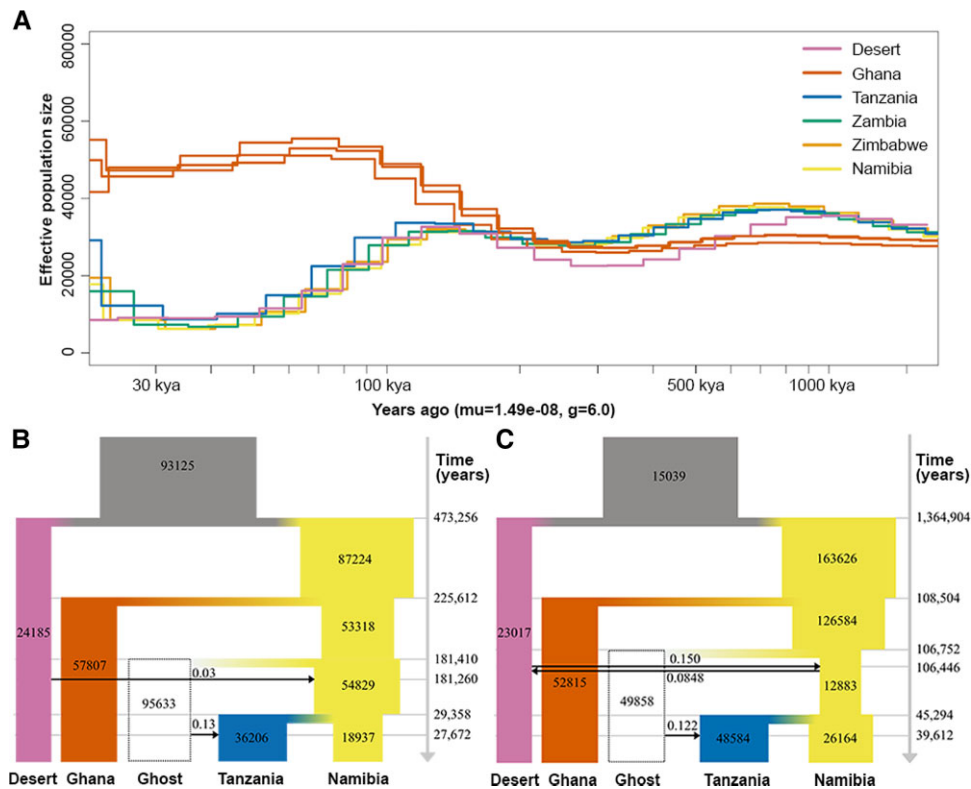
FIG. 4. Demographic history of desert warthogs and common warthogs. (*A*) Effective population size of common warthog and desert warthog populations changes across time, estimated from high-depth samples using PSMC. (*B*) Schematic diagram depicting the fastsimcoal2 demographic model and parameter point estimates, when using the best-fitting qpgraph to fix the topology of the admixture graph and admixture proportions (1-admixture model). All inferred demographic parameters and 95% confidence intervals are shown in supplementary material table S5, Supplementary Material online. (*C*) Schematic diagram depicting a more general demographic model, where the admixture proportions are estimated parameters and bidirectional migration involving common warthogs to desert warthogs introgression is allowed (3-admixture model). All inferred demographic parameters and 95% confidence intervals are shown in supplementary material table S6, Supplementary Material online.

material fig. S7, Supplementary Material online), we could not discard bidirectional gene flow to have occurred. For this reason, we decided to explore an additional demographic model with fastsimcoal2 that allows for bidirectional and asymmetric introgression between the two species, that we called "3-admixture" (fig. 4C, supplementary material fig. S9, Supplementary Material online). This resulted in estimated admixture proportions of 15% from desert warthog to ancestral ESA common warthog and 8% in the opposite direction. Based on this model, the estimated desert-common divergence time is significantly older (1,364 kya; 95% CI 1,023–1,683 kya; supplementary material table S6, Supplementary Material online). The split between the western African and ESA common warthog lineages was estimated to be 226 kya (95% CI 193–260 kya) or 108 kya (95% CI 87–165 kya), assuming unidirectional or bidirectional gene flow, respectively. The most recent divergence between eastern (Tanzania) and southern (Namibia) populations of common warthog was 29 kya (95% CI 16–44 kya) or 45 kya (95% CI 31–51 kya), respectively. Estimates of current effective population sizes for each population under either model closely resembled the results from PSMC, with the highest N_e inferred in Ghana and the lowest N_e in the desert warthog and common warthogs from Namibia (supplementary material tables S5 and S6, Supplementary

Material online). The inferred ancestral population sizes were, however, more variable across methods.

The 3-admixture model has a considerably better likelihood than the 1-admixture model, even when allowing the admixture proportions of the latter to be estimated parameters instead of fixing them (supplementary material table S4, Supplementary Material online). Due to limitations on the use of composite likelihoods for model selection, however, and the sensitivity of SFS-based divergence time estimates to demographic model assumptions, we consider both model estimates as plausible for describing the divergence history of *Phacochoerus*. Moreover, we corroborated, by estimating the population pairwise F_{ST} resulting from each of these two models, that both of them can capture the basic characteristics of the population structure in warthogs (supplementary material table S7, Supplementary Material online). Finally, we complemented the SFS-based fastsimcoal2 divergence time estimates with the TT method. This is based on single samples from each population and makes different demographic model assumptions than fastsimcoal2. This method, likewise, suggested an older divergence time of 1.2–1.3 million years between desert and common warthogs (supplementary material table S8, Supplementary Material online). To reflect the uncertainty associated

with the demographic modeling, we conservatively conclude that the species divergence time is 400–1,700 kya. Of note, these divergence times are based on the most reasonable available mutation rate for suids (see Materials and Methods).

Adaptive Introgression Scan

We used genotype calls from the high-depth samples to estimate regions of putative adaptive introgression between desert warthog and ESA common warthog populations. We used the f_d statistic, which estimates the fraction of the genome shared with a putatively introgressed ancestry, and is suited for local analyses and to detect adaptive introgression by an outlier based approach. A Manhattan plot and histogram of the genome-wide distribution show outliers of high f_d values (figs. 5A and B). The signals in the top 0.01% f_d windows are shown in supplementary material table S4, Supplementary Material online. Most of these windows also show exceptionally low values of F_{ST} between desert warthogs and ESA common warthogs compared with the genomic averages (fig. 5C, supplementary material fig. S5, Supplementary Material online). Local genomic analyses are more sensitive to mapping and genotyping errors. We tried to alleviate this problem by filtering out all genomic windows that showed an excess of problematic sites, based on the proportion of sites filtered by the reference genome filtering (see Materials and Methods). Although this does not definitely exclude all local biases, we consider the top outlier windows to be the best candidates for adaptive introgression that can be identified with the available resources. This makes them worthy of exploration here and of validation in future studies. We did not pursue simulations to obtain P-values of our selection peaks under neutrality. This was because of the many assumptions needed to accurately simulate both the population demography and the sequencing process, which we believe would make such P-values difficult to interpret. Instead, we chose to only explore the top outliers from our scan.

Nine of the 22 top outlier windows contain prominent immune system–related genes and gene clusters, including four windows in the major histocompatibility complex (MHC). In addition, the Fc gamma receptor (FCGR) locus was also among the top regions, as well as genes Mx1, Mx2, PTGS2, JAKMIP1, and ST3GAL1 (supplementary material table S4 and fig. S4, Supplementary Material online), all of which have well-characterized immune system–related functions. The MHC is well-known for its role in the adaptive immune system and, thereby, in pathogen resistance (Hill 1998). The FCGRs bind to immunoglobulin G and are hence important modulators of the immune response (Nimmerjahn and Ravetch 2006). PTGS2 encodes cyclooxygenase-2 (COX-2), which is essential for synthesizing prostaglandins, contributing to the innate immune system and inflammatory reactions (Ricciotti and FitzGerald 2011; Sander et al. 2017). The deadly African swine fever virus evades the pig immune response partly by inhibiting the expression of COX-2 (Granja et al. 2009). The two myxovirus-resistance genes (Mx1 and Mx2) are involved in resistance against viruses in general (Verhelst et al. 2013) and classical swine fever (Zhou et al. 2018). JAKMIP1 is involved in T-cell differentiation, specifically in the differentiation of virus-specific memory T cells and, therefore, in the adaptive immune system (Libri et al. 2008). ST3GAL1 plays a role both in T- and B-cell differentiation (Giovannone et al. 2018). Host variants in this gene are associated with the severity of influenza A infection (Maestri et al. 2015). Another gene in the outlier regions, MUC19, encodes a secreted mucin which is the major gel-forming mucin in pig saliva (Chen et al. 2004), and is perhaps also involved in immune system functions (Hasnain et al. 2013; McBride et al. 2018).

Discussion

A Revision of Warthog Phylogeography and its Broader Implications

Here we provide the first genome-level analyses of *Phacochoerus* and detailed insight into its previously contentious evolutionary history. We did not find support for three continental refugia during the Pleistocene (Muwanika et al. 2003), nor for a colonization of eastern Africa from southern Africa (Lorenzen et al. 2012). Instead, we found consistent evidence that extant common warthog originated in western Africa, followed by an expansion circumnavigating the central African rainforest, hence first colonizing eastern Africa and later southern Africa.

Our results reject three common warthog subspecies (*P. a. africanus*, *P. a. massaicus*, and *P. a. sundevallii*) across the sampled populations, as there is very low differentiation and a shared demographic history until the late Pleistocene between eastern and southern populations. In contrast, the Ghanaian population, representing *P. a. africanus*, is highly differentiated from ESA population, diverged long ago (108–226 kya), and has a markedly different demographic history with a much higher effective population size. Together with the exclusive desert warthog gene flow into ESA common warthogs, these differences might warrant the distinction of western and ESA common warthog populations as different subspecies or Evolutionary Significant Units (Moritz 1994). The large sampling gap between Ghana and East Africa, however, prevents us from making firm conclusions about whether the Ghana population represents a highly distinct population or the edge of a continuum.

Although our results revise several of the conclusions of previous warthog studies (Muwanika et al. 2003; Lorenzen et al. 2012) using lower numbers of genetic markers, they support some key biogeographical hypotheses. First, they show that major dispersal events in African savanna-adapted mammals alternate between an east–west axis in the northern savanna bioregion and a north–south axis in the southern savanna bioregion—the inverted L-shape coined by Kingdon (2013). Second, they corroborate East Africa as the intersection of these two savanna bioregions, serving as a "melting pot" of secondary contact
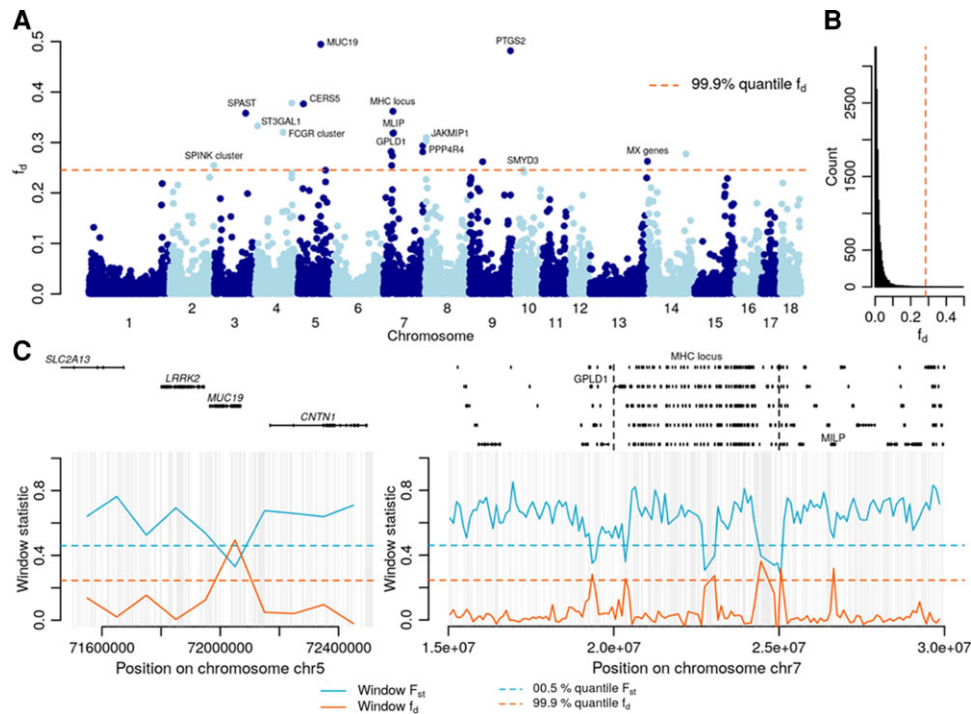
FIG. 5. Adaptive introgression scan. (*A*) Manhattan plot of f_d values estimated for 100 kb windows, estimated from called genotypes using the four high-depth common warthog from eastern and southern Africa samples with desert warthog ancestry as P2, the Ghana common warthog as P1, the desert warthog sample as P3, and the domestic pig as out-group. For windows with high desert warthog admixture, the names of the genes overlapping them are shown. (*B*) Distribution of the f_d values plotted in *A*, indicating the 99.9% quantile used as a threshold to detect outliers. (*C*) Plot of f_d and F_{ST} within two outlying F_{ST} windows and its surrounding genomic region, together with its annotated protein coding genes. Light gray-shaded areas indicate sites excluded from the analyses. Supplementary material fig. S10, Supplementary Material online shows local context plots for all outlying windows in the genome scan.

between previously vicariant lineages, as well as a mosaic refugium for species (Lorenzen et al. 2012). Herein, we provide the most detailed phylogeographic validation of these two cornerstones of African biogeography, and demonstrate that adaptive introgression may be a previously underappreciated feature within the East African contact zone. This has wide-ranging implications for understanding the biogeographic role played by eastern Africa, including that of early hominins, of which at least three species coinhabited the region during the early Pleistocene (Antón et al. 2014).

Reconciling the Fossil and Genetic Records

Our demographic modeling is inconsistent with the previously suggested molecular divergence time of 4.4–8.8 Ma (Randi et al. 2002; Gongora et al. 2011). We estimate a much more recent divergence time that ranges within 390–624 or 840–1,728 kya, assuming either unidirectional or asymmetric bidirectional gene flow between ancestral desert warthog and common warthog populations. A species divergence time within the 400–1,700 kya interval reconciles the genetic evidence with the fossil record, which finds reliable evidence of desert warthog and common warthog coexisting only since around 400 kya (Cooke and Wilkinson 1978). The earliest fossils attributed to *Phacochoerus* have been dated at 2.2–2.0 Ma (Pickford 2013a, b; Pickford and Gommery 2016). Moreover, our

inferred divergence time refutes the possibility that the two warthog species diverged in Eurasia before moving to Africa (Randi et al. 2002; Gongora et al. 2011), leaving no fossil traces for millions of years in Africa. The inconsistency between the genetic and fossil sources of evidence has obscured the evolutionary relationship between the two extant species of warthog, for example, by leading researchers to assume that any interbreeding between the two taxa was highly unlikely (De Jong and Butynski 2018).

The discrepancy between our coalescent-based estimate of divergence time and previous phylogenetic estimates can probably be attributed to two factors. First, genomic divergence predates species divergence by an expected $2N_e$ generations due to the presence of polymorphism in the ancestral population (Edwards and Beerli 2000), but this distinction can only explain a minor proportion of the discrepancy. Second, most of the literature (e.g., Frantz et al. 2013, 2015, 2016; Groenen 2016) used node calibrations that originated in the first phylogenetic studies based on mtDNA and a few nuclear makers (Randi et al. 2002; Gongora et al. 2011). Any early overestimations of lineage split times would, therefore, have been propagated in subsequent studies. A recent study highlights the possibility that suid evolutionary rates were highly overestimated (Zhang et al. 2022). Collectively, these results demonstrate the challenges of inferring mutation rates that are accurate over evolutionary time scales, and

the effect of mutation rate uncertainty on the dating of evolutionary events.

Adaptive Introgression of Disease Resistance

We show that the ESA lineage of common warthogs admixed with desert warthogs after splitting from the western African lineage 87–260 kya. The majority of this gene flow is inferred to be from desert warthogs to the ESA common warthog lineage. We hypothesized that—similar to Neanderthal (*Homo neanderthalensis*) introgression into humans (Sankararaman et al. 2014)—desert warthogs introgression contributed genetic variation beneficial to the ESA common warthogs. The Zambezian (eastern and southern African) savanna biomes differ from the Sudanian (western) savanna biome in vegetation, climate, and other aspects (Happold and Lock 2013). We, therefore, searched for candidates for adaptive introgression in ESA common warthog populations. The prevalence of immune system–related genes among the f_d outliers was striking, with the MHC locus being particularly noteworthy. Pathogen resistance has previously been identified as a major driver of positive selection on genomic segments that introgressed from Neanderthals to modern humans (Dannemann et al. 2016; Deschamps et al. 2016; Quach et al. 2016; Enard and Petrov 2018). Pathogens are a major driver of selection in many species, perhaps particularly when a species expands its range and encounters exotic pathogens to which it has no pre-existing resistance (Karlsson et al. 2014). In such cases, adaptive introgression of resistance-conferring variants from a closely related native species is a plausible evolutionary scenario compared with selection on standing variation or de novo mutations (Hedrick 2013). It is likely that common warthogs encountered new pathogens as they moved eastwards and southwards through Africa, whereas desert warthogs may have had a head start of hundreds of thousands of years in adapting to such pathogens. One well-known and highly lethal suid pathogen naturally endemic to eastern and southern Africa, but not western Africa (Taylor 1977; Costard et al. 2009; Jori et al. 2013; Zhu et al. 2019), is African swine fever virus (Zhu et al. 2019; Wang et al. 2020). Intriguingly, three genes identified as adaptively introgressed in our study—the two *Mx* genes and *PTGS2*—play key roles in the immune response to African and classical swine fever viruses (Netherton et al. 2009; Zhou et al. 2018; Fan et al. 2020). This raises the possibility that African swine fever, or a similarly deadly pathogen, drove adaptive introgression in warthogs.

The signal of ancient introgression, together with our estimate of a shorter time of divergence, opens up the possibility that hybridization between the two warthog species can still occur. Hybridization would be important from a species conservation perspective. Although we did not detect hybrids, one common warthog from an area of sympatry in Kenya showed signs of increased desert warthog ancestry. Based on our findings, rare but on-going hybridization cannot be excluded, especially where the two species are broadly sympatric (Butynski and De Jong 2021; De Jong et al. in press).

Conclusions

We solve the long-standing riddle of the time of divergence of the two extant species of warthog and show how an eastward range expansion in common warthogs brought common warthog and desert warthog into contact. We found evidence of introgression between the two species. This occurred either in the Sudanian savanna region or in eastern Africa. Our results suggest that pathogen resistance played a prominent role in driving introgressed genomic segments from desert warthogs to high frequencies in eastern and southern common warthogs.

Materials and Methods

Sample Collection and Laboratory Protocol

Fifty-five samples of warthog tissue, mainly dried skin, were used in this study (supplementary material table S1, Supplementary Material online). They were collected during 1994–1999 in Ghana, Kenya, Tanzania, Zambia, Zimbabwe, and Namibia, and most were previously analyzed by Muwanika et al. (2003) who describe sample collection and storage. Six of the samples are from desert warthogs. The samples from common warthog cover three of the four currently recognized subspecies of common warthog. We also included samples from a giant forest hog (*Hylochoerus meinertzhageni*), a red river hog (*Potamochoerus porcus*), a bush pig (*Po. larvatus*), and two domestic pigs (*Sus scrofa*) from Africa (supplementary material table S1, Supplementary Material online). Following the manufacturer's protocol, the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA, USA) was used for DNA extraction. Subsequently, RNase was added to the samples to ensure they consist of RNA-free genomic DNA. DNA concentrations were then measured with a Qubit 2.0 Fluorometer and a Nanodrop before using gel electrophoresis to check the quality of the genomic DNA.

Sequencing

All samples were sequenced using Illumina paired-end 150 bp reads. Forty-nine were sequenced to low depth (about 2–5X depth of coverage) on the Illumina NovaSeq platform. Ten samples were sequenced to medium-high depth (about 15–20X) on the Illumina HiSeq2500 platform. A total of 9.07 billion raw reads were generated and analyzed. Before mapping we assessed the quality of the raw reads using FastQC (Andrews 2010) and MultiQC (Ewels et al. 2016). We also downloaded publicly available whole-genome sequencing data of a babirusa (*Babyrousa babyrussa*) (Liu et al. 2019).

Processing and Mapping of Sequencing Data

Prior to mapping, we processed the paired-end reads with NGmerge (Gaspar 2018) to merge all read pairs that overlapped by at least 11 bp. We then mapped nonmerged paired-end reads and merged single-end reads with bwa mem v0.7.17 (Li and Durbin 2010) separately in paired-end

and single-end mode, using default settings and mapping to the domestic pig genome assembly Sscrofa 11.1. We marked and removed duplicate reads with samtools v. 1.9 (Li et al. 2009), and removed other low quality alignments using the samtools flag -F 3852 and nonproperly paired reads with samtools flag -f 3. Finally, we used samtools to combine nonmerged and merged aligned reads for each sample into a single BAM file.

Data Filtering

Throughout all analyses we excluded bases with a base call quality below 30, as well as reads with a mapping quality below 30. Furthermore, unless otherwise stated, we restricted all analyses to autosomal chromosomes and excluded regions annotated as repeats in the reference genome (Sscrofa 11.1). In addition, we conducted a series of sample-filtering steps and site-filtering steps that we explain in the following sections.

Sample Filtering

We first filtered the samples to exclude those with excessive per base sequencing error rates as well as closely related or duplicate individuals.

Error rate Estimation

We used the "perfect-individual" method described in Orlando et al. (2013) and incorporated in ANGSD (Korneliussen et al. 2014) to estimate the per base sequencing error rate for each individual. This method measures error rates as excess mismatches between each sample and the out-group (domestic pig reference), relative to the mismatches between the out-group and the "perfect-individual." This excess corresponds to the per base sequencing error rate assuming that all individuals are equally distant to the out-group and that the consensus sequence from the "perfect-individual" has few errors. Sample 1257 was chosen as the "perfect-individual" because it has higher depth. This sample can, therefore, be used to create a consensus sequence with few errors. This was done using ANGSD (-doFasta 2) and strict quality filters, including a minimum base quality of 35, minimum mapping quality of 35, minimum sequencing depth per site of 10, and keeping only uniquely mapping reads. Per base sequencing error rates were then estimated using all bases (-doAncError 1). Based on the results, we excluded sample ID 6436 due to excessive errors (supplementary material fig. S11, Supplementary Material online). After all site filters were applied (see below), we re-estimated the per-sample error rates using the same approach but only considering the sites qualified after site filtering (see below). This was used to correlate sample heterozygosity with sample error rate (supplementary material fig. S5, Supplementary Material online).

Identification and Removal of Close Relatives

We used the allele frequency–free method described in Waples et al. (2019) to identify duplicate or closely related samples based on the R0, R1, and KING-robust statistics.

These statistics are based on identity by state (IBS) between pairs of samples. We estimated GLs with ANGSD for all desert warthog and common warthogs jointly, using the GATK model (-GL 2; McKenna et al. 2010), and called single nucleotide polymorphisms (SNPs) using a P-value threshold of 10^{-6} and a MAF filter of 0.05. We then used the GLs as input for NGSrelate (Hanghøj et al. 2019; Waples et al. 2019), which implements the estimation of the three IBS statistics. We identified three groups of sample duplicates. The duplicated samples, all of which came from Ghana, had KING-robust kinship values >0.46, R1 values >6.37, and R0 values $<2 \times 10^{-6}$. We excluded all but one sample from each duplicated group, retaining samples 7152, 7155, and 6274 (supplementary material table S10, Supplementary Material online). For some analyses, we pooled the sequencing data for each duplicated individual to obtain higher depth. We also identified five pairs of first-degree relatives and removed one sample from each pair (supplementary material table S10, Supplementary Material online). The final data set for analysis consisted of four desert warthogs and 35 common warthogs (supplementary material tables S1 and S10, Supplementary Material online).

Site and Reference Sequence Filtering

We performed a series of quality controls on the reference sequence, as ambiguous regions can impact downstream analyses (Pečnerová et al. 2021). For these site-filtering steps, we used only the samples retained after the sample-filtering steps outlined above.

Mappability Filter

We estimated mappability with GENMAP (v1.2.0; Pockrandt et al. 2020) conservatively using 100 bp k-mers with up to two mismatches (and otherwise default settings) to compute the mappability scores for each site. Consequently, sites with a score <1 were excluded from further analyses.

Global Depth Filter

We estimated global depth per site separately across low- and high-depth desert warthog and common warthog samples using ANGSD (Korneliussen et al. 2014). We then estimated the median depth per site across each of the low- and high-depth data sets, excluding sites that had a depth below half the median or above 1.5 times the median for any of the two sets.

Excess Heterozygosity (Hardy–Weinberg equilibrium) Filter

We identified regions with excess of heterozygosity, which is indicative of mapping problems, using the Hardy–Weinberg equilibrium (HWE) likelihood ratio test built into the PCAngsd framework (Meisner and Albrechtsen 2018, 2019). This accounts for population structure in estimating per site inbreeding coefficients. Inbreeding coefficients (F) take values in the range from −1, when there is a total excess of heterozygous genotypes, to 1, indicating a total excess of homozygous genotypes, whereas 0 corresponds to having genotype in HWE proportions within

each ancestry. It can, therefore, be used to identify regions with a strong excess of heterozygosity.

We first used ANGSD to estimate GLs for all common warthogs that passed sample quality control. We did not use any site filtering, except basic base quality and mapping quality filters, calling SNPs with MAF $\geq$0.05 and SNP P-value $<10^{-6}$ and using the GATK GL model implemented in ANGSD. We used the GLs as input for PCAngsd, using the first three principal components (PCs) to obtain the individual allele frequencies used to correct for population structure (Meisner and Albrechtsen 2019). We considered sites with $F < -0.95$ and P-value $<10^{-6}$ to be exclusively heterozygous in the ancestral populations that are polymorphic and excluded all sites within 10 kb of such sites.

All the site filters were used in subsequent analyses unless stated otherwise.

Data Analyses
GLs, SNPs, and Genotype Calling
We used GLs or single read sampling for all analyses in which the low-depth samples were analyzed in order to account for the genotype uncertainty of calling genotypes (da Fonseca et al. 2016).

ANGSD was used to estimate the GLs by applying the GATK model (-GL 2), inferring the major and minor allele from the GLs (-doMajorMinor 1), estimating the allele frequencies (-doMaf 1) and, where applicable, calling SNPs using the default likelihood ratio test (-SNP_pval 1e-6) and a minimum allele frequency filter of 0.05.

We called genotypes for the high-depth samples using bcftools v. 1.10 consensus caller (-c) (Li et al. 2009), and used bcftools v1.10 throughout to manipulate the called genotype files (Danecek et al. 2021). We called genotypes per sample for all sites, both variable and fixed, and using only the base quality and read mapping quality filters. Further filtering of sites depended on the analyses and is described in the section corresponding to each of those analyses.

mtDNA Analyses
To check the taxonomic status of our samples, we initially performed a mitochondrial genome mapping and consensus calling for each of our warthog and out-group samples. We also included the published mtDNA sequence of a domestic pig, a red river hog, and a babirusa. To generate the consensus sequences, we mapped all desert warthog and common warthog samples, and the giant forest hog sample to the common warthogs mitochondrial genome (NC_008830; Wu et al. 2007), the bush pig sample to the red river hog mitochondrial genome (NC_020737; Hassanin et al. 2012), and the babirusa to the domestic pig mitochondrial genome (NC_000845; Lin et al. 1999). In all cases, we used the same pre- and post-processing of reads as with the nuclear genome mapping. We used the mapped data to build consensus mtDNA sequences for each sample using ANGSD (Korneliussen et al. 2014), by keeping the consensus base in

each position (-doFasta 2). We removed heteroplasmic sites by masking (as "N") any mitochondrial site where <95% of reads carried the same base.

After consensus calling, we aligned the mitochondrial genomes using BioEdit (Hall et al. 2011) and used the alignment in a BEASTv.1.8.4 (Drummond and Rambaut 2007) phylogenetic analysis. We used a GTR + G + I model and a coalescent BSP prior to avoid restricting the tree prior by imposing a narrow population size range. The population size prior was set to a uniform range between 10^3 and 10^6 to reflect that we use it as a nuisance parameter. Default priors were used for the other parameters. We used a single node calibration prior on the time of the most recent common ancestor of Suinae (all individuals except the babirusa). The prior was normally distributed with mean 10^7 years and standard deviation 10^6 years based on Frantz et al. (2016). We ran the MCMC chain for 10^7 steps, sampling trees, and parameters every 1000 steps. Convergence and proper mixing were assessed by visual inspection and by estimating parameter ESSs using TRACER (Rambaut et al. 2018). We used TreeAnnotator (Helfrich et al. 2018) to make a Maximum Clade Credibility (MCC) tree, discarding the first 1000 trees. Finally, we used iTol (Letunic and Bork 2021) to visualize the MCC tree together with node posteriors and a time scale.

Population Structure
Principal Component Analysis
We used PCAngsd (Skotte et al. 2013; Meisner and Albrechtsen 2018) to estimate the genotype covariance matrix of the 35 common warthogs left after sample filtering. We used two PCs to estimate the individual allele frequencies used to estimate the covariance matrix. This was detected as the optimal number of PCs to model the population structure based on Velicer's minimum average partial test implemented in PCangsd.

Analyzing and Evaluating Population Admixture
We estimated admixture proportions for 35 common warthogs using NGSadmix (Skotte et al. 2013) based on GLs. We ran NGSadmix from $K = 2$ to $K = 7$ until either the results converged, which we defined as a maximum difference of two log-likelihood units between the top three maximum likelihood results, or 100 independent runs finished without convergence. For the runs that converged, we subsequently evaluated the model fit using evalAdmix (Garcia-Erill and Albrechtsen 2020). A positive correlation of the pairwise residuals indicates a poor model fit and can be used to identify the best-fitting value of K.

EEMS
We used ANGSD to generate an IBS matrix for 34 common warthogs for which we had sample location data (excluding the Zambian sample) by sampling a single read at each SNP (called with -snp_pvaL 1e6 and -minMaf 0.05) where we had data for both individuals (-doIBS 1 -makeMatrix 1).

Thus, the distance between samples is calculated at each site (0 or 1) and averaged over all used sites.

This matrix was used as input for Estimated Effective Migration Surfaces (EEMS) (Petkova et al. 2016) along with coordinates of the sample origin, and analyzed using runeems_snps. We used 10 million steps and a burn-in of 2 million steps with 400 demes and 10,176,836 sites in total. The results were visualized using rEEMSplots.

Population Differentiation

We used ANGSD to generate per-population site allele frequency (saf) files from GLs for every population with more than one sample. For the Kenyan population, we excluded the sample modeled as a mixture of different clusters in NGSadmix. For each pairwise population, we then used realSFS (Nielsen et al. 2012) to estimate population pairwise 2dSFS from saf files based on randomly sampled blocks of contiguous nonfiltered sites until reaching 200 Mbp. We then used these 2dSFS as a prior for estimating genome-wide F_{ST} from saf files estimated from all autosomal sites (except those removed by the previously described site filters), using Hudson's F_{ST} estimator (Bhatia et al. 2013). We also estimated F_{ST} between pairs of single high-depth samples, estimated from the 2dSFS between each pair of individuals using the called genotypes as input. This has been shown to be accurate in similar situations (Pečnerová et al. 2021).

Directionality of Range Expansion

To corroborate the direction of the range expansion, we estimated a directionality index among common warthog populations (Peter and Slatkin 2013). We included in this analysis those populations with a high-depth sample and known coordinate information of the sampling locality. This resulted in keeping four populations, each with one sample. We based the analyses on called genotypes, applying the same filters as when estimating D-statistics from called genotypes (see below) and keeping only sites variable within the four common warthog samples. The red river hog sample as an out-group to assign ancestral and derived states to each SNP. We then used the R package "rangeexpansion" to infer the directionality index psi between each population pair.

Heterozygosity

We estimated the genome-wide heterozygosity for all samples using ANGSD and realSFS by estimating the individual SFS and then dividing the number of heterozygous sites by the total number of sites. For the six high-depth samples, we also estimated the genome-wide heterozygosities from the genotypes called with bcftools by counting the total number of heterozygous sites of each sample and dividing by the total number of sites remaining after filtering.

D-Statistics

We used D-statistics (Green et al. 2010) to investigate presence of gene flow between desert warthogs and common warthog populations, and to investigate potential

introgression of an out-group to both warthog species that might confound the analyses. We estimated D-statistics for both low- and high-depth samples, using different approaches for each.

For the low-depth samples, we applied the single read sampling approach implemented in ANGSD (--doAbbaBaba 1). Only sites that passed all reference genome filters, with a domestic pig (sample ID 1003) as out-group, were used. We used a block jackknife approach to estimate standard errors and calculate Z-scores (Busing et al. 1999), using blocks of 5 Mbp size. We also estimated D-statistics based on single read sampling for a subset of common warthogs from Tanzania and Ghana, with the domestic pig samples as P3 and a babirusa as the out-group. This was to investigate potential confounding gene flow of an out-group of all extant African suids into the Ghanian common warthog population.

For the high-depth samples, we estimated D-statistics based on called genotypes. We used bcftools to filter the called genotypes with the following criteria. We removed sites with mappability below one and sites within annotated repeats in the reference genome as in other analyses. Due to the inclusion of out-group samples that had not been considered in the filters based on depth and excess heterozygosity, we did not use the previous site filters. Instead, we filtered based on the set of genotype calls, removing sites where any of the samples had a depth below 10 or above 50, kept only biallelic SNPs, removed sites where all samples were called as heterozygous, and removed sites where any sample had a heterozygous call with less than three reads supporting either of the two alleles. Based on the called genotypes, we used the qpDstat program from the package AdmixTools (Patterson et al. 2012) to calculate the counts of ABBA and BABA sites for trees where desert warthog was P3, domestic pig was the out-group, and with all possible combinations of common warthog populations as P1 and P2. We furthermore estimated D-statistics with the Ghanaian common warthog population as P1, the desert warthog or Tanzanian common warthog as P2, the red river hog sample as P3 and the domestic pig as out-group. This was done to investigate potential gene flow of an out-group of the *Phacochoerus* genus to the Ghanaian common warthog that could confound the results. We applied block jackknife to estimate standard errors and calculate Z-scores, using 461 blocks of 5 cM and assuming a uniform conversion of 1 cM/Mbp.

Treemix and qpGraph

We ran TreeMix (Pickrell and Pritchard 2012) on 34 common warthogs (excluding the Kenyan sample modeled as admixed in the NGSadmixe analyses) grouped by the country of origin of the samples and on four desert warthogs and included the two domestic pig samples as an out-group. We generated the input allele counts per population by first estimating the likelihood of sample allele frequency for each population with ANGSD. We used only sites that passed all site filters, polarized using the

reference domestic pig genome as the ancestral state. We then merged all populations by keeping only sites where all populations had data, and called within each population the maximum likelihood sample frequency as the allele count. This procedure resulted in approximately 17 million sites as input data. For each possible migration from 0 to 3, we ran 25 differently seeded replicates of TreeMix and chose the run with the highest likelihood, breaking ties at random. We set domestic pig as an out-group, and used a block size of 28 k input sites, corresponding to an assumed maximum LD of 5 Mbp.

We ran qpGraph from the package AdmixTools (Patterson et al. 2012) on the high-depth individuals from a subset of the populations (Desert, Ghana, Tanzania, and Namibia) using the heuristic graph search tool qpBrute (Ní Leathlobhair et al. 2018; Liu et al. 2019). We used the same set of genotype calls as used to estimate the D-statistics from high-depth samples. Z-scores were estimated with block jackknife using blocks of 5 Mbp. We ran qpGraph on all SNPs from the filtered sites and, otherwise, used the default settings.

Adaptive Introgression Scan

We scanned the genome of the common warthog samples from eastern and southern Africa for regions showing strong signals of introgression from desert warthogs. We based this analysis on the same genotype call set as used for estimating D-statistics (see above). We estimated local admixture proportions between pooled high-depth ESA samples (i.e., we pooled the four high-depth samples from Tanzania, Zambia, Zimbabwe, and Namibia) and the high-depth desert warthog sample using the ABBA-BABA based f_d statistic (Martin et al. 2015) in windows of 100 kb. The f_d statistic measures the fraction of the genome shared due to introgression between populations P2 and P3. This is measured as the fraction of excess sharing of derived alleles between P2 and P3 relative to between a P1 and the P3 population. This P1 population must be sister to P2 and assumed to not have received any introgression. An out-group is used to polarize derived alleles. This statistic allows for the possibility of bidirectional introgression between P2 and P3 (Martin et al. 2015). It has been shown that this statistic is powerful for detecting adaptive introgression when selection is intermediate or strong (Racimo et al. 2017). We used the high-depth sample from Ghana as P1 and a domestic pig (sample ID 1003) as an out-group. We also estimated F_{ST} in sliding windows between desert warthogs and the pooled ESA common warthog samples using both low- and high-depth samples. We estimated F_{ST} in genomics windows with ANGSD by estimating safs for both populations, and then using these saf files d to estimate a 2dSFS. The saf file was then also used to estimate F_{ST} in 100 kb windows, using the Hudson estimator and with the 2dSFS as prior (Korneliussen et al. 2013).

We restricted these sliding window f_d and F_{ST} analyses to sites retained after the reference genome filtering outlined above. Furthermore, we removed any windows in the bottom 5% of missing data, which resulted in keeping windows with information in at least 34.2% of sites. We selected the top 0.1% of the remaining f_d values and lumped all outlier windows within 1 Mbp of each other into a single signal. We considered these regions as candidates of adaptive introgression and extracted the genes contained in these windows from the domestic pig reference genome annotation. We acknowledge that our adaptive introgression scan is sensitive to the relatively small sample sizes, as well as possibly biased by mapping issues given that we map our data to the distantly related domestic pig genome. We addressed the latter bias by imposing a strict set of filters for the inclusion of sites in the analyses (see above).

Demographic History Inference

Mutation Rates and Dating

Previous studies used a rate of 2.5×10^{-8} mutations per site per generation for demographic analyses in suids (Groenen et al. 2012). This was based on now-obsolete high estimates of the human mutation rate. This appears excessively high compared with rates inferred from phylogenomic analyses in ruminants (Chen et al. 2019). Given that mutation rates are notoriously difficult to estimate, in our main results, we used the mean rate of 2.48×10^{-9} mutations per site per year phylogenetically estimated across wild ruminants (Chen et al. 2019), which have life histories comparable with warthogs. This rate is, furthermore, almost equal to the mean annual rate of 2.2×10^{-9} mutations per site per year inferred across a broad range of mammals (Kumar and Subramanian 2002). We assumed a generation time of 6 years in accordance with Pacifici et al. (2013) and therefore a per generation mutation rate of 1.49×10^{-8}. The consensually agreed mutation rate for cattle is 1.17×10^{-8} per generation estimated using pedigrees and a reproductive age of 5 years (Harland et al. 2017). This, again, is similar to the rate we assumed for warthogs. Recently, Zhang et al. (2022) used domestic pig pedigrees to estimate a much lower mutation rate of 3.6×10^{-9} per site per generation, but mutation rates estimated through pedigrees are very sensitive to filtering choices as they depend on the ability to successfully distinguish between low numbers of true de novo mutations and sequencing errors or other artifacts (Bergeron et al. 2022). We note that inferred dates scale about linearly with the assumed mutation rate, so our dating estimates can be readily converted should more accurate estimates of the suid mutation rate become available.

PSMC

Using PSMC (Li and Durbin 2011), we estimated the effective population sizes back through time. We ran PSMC on the six high-depth–sequenced samples and added two high-depth samples from Ghana obtained by merging several duplicate low-depth samples (see above and supplementary material table S1, Supplementary Material online). We applied the reference filters and called genotypes using bcftools v1.10. We removed sites

covered by less than ten reads or more than two times each sample mean depth, as well as sites called as heterozygous where any of the two alleles was in less than three reads. We used default settings for all PSMC parameters. Results were scaled to real time using a generation time of 6 years following (Pacifici et al. 2013) and a mutation rate of 1.49×10^{-8} per generation (as discussed above).

Fastsimcoal

The demographic history of a representative subset of the common warthog populations (Ghana, Tanzania, and Namibia) and desert warthogs was further investigated using a coalescent simulation based method implemented in fastsimcoal2 v2.6.0.3 (Excoffier et al. 2013). To minimize potential bias arising when determining ancestral allelic states, we used the folded 2dSFS based on randomly subsampled 200 Mbp (see F_{ST} estimation above) as input for the inference. Five plausible demographic models were tested (supplementary material fig. S5, Supplementary Material online). As a baseline model, the "No-admixture" scenario is a model without any admixture events and follows a population tree equivalent to the TreeMix tree without any migration edges. We included a ghost population for consistency with the two other models. The 1-admixture scenario adds two admixture events with admixture proportions fixed to the results from the best-fitting qpGraph. The "2-admixture" models a scenario in which there is admixture between the extinct Cape warthog, modeled here as a ghost population branching off the desert warthog, and Namibia. We also considered a model 3-admixture similar to 1-admixture, but where admixture between desert warthogs and ancestral ESA common warthogs is bidirectional rather than unidirectional. Finally, to check whether the better fit of the bidirectional gene flow model 3-admixture relative to 1-admixture was due to not fixing the admixture proportions, we included an additional version of the 1-admixture model where the unidirectional admixture is not restricted to the values inferred by qpGraph.

For each model, we ran 50 independent runs to find the best-fitting parameters yielding the highest likelihood, with 100,000 coalescent simulations per likelihood estimation (-n100000) and 20 conditional maximization algorithm cycles (-L20). The model fits were assessed by comparing maximum likelihoods (Excoffier et al. 2013), despite limitations due to using composite likelihoods. In order to evaluate the fit of the unidirectional and bidirectional gene flow models 1-admixture and 3-admixture to the real data sets, we computed Hudsons F_{ST} based on the simulated joint SFS for the maximum likelihood parameters. To obtain the 95% CI, we generated 100 parametric bootstraps based on the maximum likelihood parameters estimated under the best model and ran 50 independent runs for each bootstrap, using the same settings as for the analyses of the original data set. A mutation rate of 1.49×10^{-8} per site per generation (see Mutation Rates and Dating) and a generation time of 6 years (Pacifici et al. 2013) were used to convert model estimates from coalescence units to absolute values (i.e., years).

Split Time Estimates with the TT Method

We applied the TT method (Schlebusch et al. 2017; Sjödin et al. 2021) as a complement and validation of the split times inferred with fastsimcoal2. This method is based on modeling the probability of the genotype combinations from two diploid samples from each population, as a function of several parameters that include divergence times. It does not make any assumption on the population sizes after the split, but it assumes a constant ancestral population size and a clean split without posterior gene flow. We obtained the genotype combinations as the 2DSFS between the high-depth desert warthog sample and each of the five high-depth common warthog samples, and polarized the ancestral state as the allele in the domestic pig reference. In addition, we excluded from the analyses those sites for which the babirusa and red river hog samples were not homozygous for the reference allele.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

The raw sequencing data generated for this project have been deposited in FastQ format in SRA with BioProject accession code PRJNA837362. Scripts for data analyses were developed using R (R Core Team 2018), python3 (Van Rossum and Drake 2009), and snakemake (Mölder et al. 2021) and can be found in the github page https://github.com/GenisGE/warthogscripts.

References

FAO. 2012. *Global Ecological Zones for FAO Forest Reporting: 2010 Update*. Rome, Italy: Food and Agriculture Organization of the United Nations.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data [cited 2022 Jun 21]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antón SC, Potts R, Aiello LC. 2014. Evolution of early *Homo*: an integrated biological perspective. *Science*. **345**:1236828.

Arctander P, Johansen C, Coutellec-Vreto MA. 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Mol Biol Evol*. **16**:1724–1739.

Bergeron LA, Besenbacher S, Turner T, Versoza CJ, Wang RJ, Price AL, Armstrong E, Riera M, Carlson J, Chen H-Y, *et al*. 2022. The Mutationathon highlights the importance of reaching standardization in estimates of pedigree-based germline mutation rates. *Elife*. **11**:e73577.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Res*. **23**:1514–1521.

Busing FMTA, Meijer E, Van Der Leeden R. 1999. Delete-m jackknife for unequal m. *Stat Comput*. **9**:3–8.

Butynski TM, De Jong YA. 2018. Common warthog (*Phacochoerus africanus*). In: Melletti M, Meijaard E, editors. *Ecology, conservation and management of wild pigs and peccaries*. Cambridge, UK: Cambridge University Press. p. 85–100.

Butynski TM, De Jong YA. 2021. Sympatry between desert warthog *Phacochoerus aethiopicus* and common warthog *Phacochoerus africanus* in Kenya, with particular reference to Laikipia County. *Suiform Soundings*. **20**:33–44.

Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, *et al*. 2019. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* **364**-(6446):eaav6202.

Chen Y, Zhao YH, Kalaslavadi TB, Hamati E, Nehrke K, Le AD, Ann DK, Wu R. 2004. Genome-wide search and identification of a novel gel-forming mucin *MUC19/Muc19* in glandular tissues. *Am J Respir Cell Mol Biol*. **30**:155–165.

Cooke HBS. 1994. *Phacochoerus modestus* from Sterkfontein Member-5. *S Afr J Sci*. **90**:99–100.

Cooke HBS, Wilkinson AF. 1978. Suidae and Tayassuidae. In: Maglio VJ, Cooke HBS, editors. *Evolution of African Mammals*. Cambridge, MA: Harvard University Press. p. 435–482.

Costard S, Wieland B, de Glanville W, Jori F, Rowlands R, Vosloo W, Roger F, Pfeiffer DU, Dixon LK. 2009. African swine fever: how can global spread be prevented? *Philos Trans R Soc Lond B Biol Sci*. **364**:2683–2696.

Cumming DHM. 2013. *Phacochoerus africanus*, common warthog. In: Kingdon J, Hoffman M, editors. *Mammals of Africa*. London (GB): Bloosmbury. Volume 6: Pigs, hippopotamuses, chevrotain, giraffes, deer and bovids. p. 54–60.

da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. 2016. Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Mar Genomics*. **30**:3–13.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*. **10**(2): giab008.

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of neandertal- and denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am J Hum Genet*. **98**: 22–33.

De Jong YA, Butynski TM. 2018. Desert warthog (*Phacochoerus aethiopicus*). In: Melletti M, Meijaard E, editors. *Ecology, conservation and management of wild pigs and peccaries*. Cambridge, UK: Cambridge University Press. p. 101–113.

De Jong YA, Butynski TM. 2021. New desert warthog records for Laikipia County, central Kenya [cited 2022 Jun 21]. Available from: https://www.wildsolutions.nl/desert-warthog-laikipia/.

De Jong YA, d'Huart JP, Butynski TM. in press. Biogeography and conservation of desert warthog (*Phacochoerus aethiopicus*) and common warthog (*Phacochoerus africanus*) in the Horn of Africa. Mammalia.

Deschamps M, Laval G, Fagny M, Itan Y, Abel L, Casanova J-L, Patin E, Quintana-Murci L. 2016. Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. *Am J Hum Genet*. **98**:5–21.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. **7**:214.

Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*. **54**:1839–1854.

Enard D, Petrov DA. 2018. Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. *Cell*. **175**:360–371.e13.

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

Ewer RF. 1956. The fossil suids of the transvall caves. *Proc Zool Soc Lond*. **127**:527–544.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet*. **9**:e1003905.

Fan W, Jiao P, Zhang H, Chen T, Zhou X, Qi Y, Sun L, Shang Y, Zhu H, Hu R, *et al*. 2020. Inhibition of African swine fever virus replication by porcine type I and type II interferons. *Front Microbiol*. **11**:1203.

Frantz LAF, Meijaard E, Gongora J, Haile J, Groenen MAM, Larson G. 2016. The evolution of suidae. *Ann Rev Anim Biosci*. **4**:61–85.

Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, Semiadi G, Meijaard E, Li N, Crooijmans RPMA, *et al*. 2013. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol*. **14**: R107.

Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Cagan A, Bosse M, Paudel Y, Crooijmans RPMA, Larson G, Groenen MAM. 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet*. **47**:1141–1148.

Garcia-Erill G, Albrechtsen A. 2020. Evaluation of model fit of inferred admixture proportions. *Mol Ecol Resour*. **20**:936–949.

Gaspar JM. 2018. NGmerge: Merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinform*. **19**:536.

Giovannone N, Antonopoulos A, Liang J, Geddes Sweeney J, Kudelka MR, King SL, Lee GS, Cummings RD, Dell A, Barthel SR, *et al*. 2018. Human B cell differentiation is characterized by progressive remodeling of O-linked glycans. *Front Immunol*. **9**:2857.

Gongora J, Cuddahee RE, do Nascimento FF, Palgrave CJ, Lowden S, Ho SYW, Simond D, Damayanti CS, White DJ, Tay WT, *et al*. 2011. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla): Evolution of extant African Suidae. *Zoologica Scripta* **40**:327–335.

Granja AG, Sánchez EG, Sabina P, Fresno M, Revilla Y. 2009. African swine fever virus blocks the host cell antiviral inflammatory response through a direct inhibition of PKC-theta-mediated p300 transactivation. *J Virol*. **83**:969–980.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y. 2010. A draft sequence of the Neandertal genome. *Science*. **328**:710–722.

Groenen MAM. 2016. A decade of pig genome sequencing: A window on pig domestication and evolution. *Genet Sel Evol*. **48**:23.

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, *et al*. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. **491**:393–398.

Grubb P. 1993. The Afrotropical suids (*Phacochoerus, Hylochoerus* and *Potamochoerus*). Taxonomy and description. In: Oliver WLR, editor. *Pigs, peccaries and hippos: Status survey and conservation action plan*. Gland, Switzerland: IUCN. p. 66–75.

Grubb P. 2005. Order Artiodactyla. In: Wilson DER, editor. *Mammal species of the world: A taxonomic and geographic reference*. Vol. 1. Baltimore (MD): The Johns Hopkins University Press. p. 637–722.

Grubb P, D'Huart J-P. 2013. *Phacochoerus aethiopicus* Desert Warthog. In: Kingdon J, Hoffmann M, editors. *Mammals of Africa: Volume VI: Pigs, hippopotamuses, chevrotain, giraffes, deer and bovids*. London: Bloomsbury Publishing. p. 51–53.

Hall T, Biosciences I, Carlsbad C. 2011. BioEdit: An important software for molecular biology. *GERF Bull Biosci*. **2**:60–61.

Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS. 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience*. **8**: giz034.

Happold D, Lock JM. 2013. The biotic zones of Africa. In: Kingdon J, Happold DCD, Butynski TM, Hoffman M, Happold M, Kalina J, editors. *Mammals of Africa*. London: Bloomsbury Publishing. Volume 1: Introductory chapters and Afrotheria. p. 57–74.

Harland C, Charlier C, Karim L, Cambisano N, Deckers M, Mni M, Mullaart E, Coppieters W, Georges M. 2017. Frequency of mosaicism points towards mutation-prone early cleavage cell divisions in cattle. *BioRxiv [Preprint]*. doi:10.1101/079863.

Harris JM, Cerling TE. 2002. Dietary adaptations of extant and Neogene African suids. *J Zool*. **256**:45–54.

Harris JM, White TD. 1979. Evolution of the Plio-Pleistocene African Suidae. *Trans Am Philos Soc*. **69**:1.

Hasnain SZ, Gallagher AL, Grencis RK, Thornton DJ. 2013. A new role for mucins in immunity: Insights from gastrointestinal nematode infection. *Int J Biochem Cell Biol*. **45**:364–374.

Hassanin A, Delsuc F, Ropiquet A, Hammer C, van Vuuren B J, Matthee C, Ruiz-Garcia M, Catzeflis F, Areskoug V, Nguyen TT, *et al*. 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol*. **335**:32–50.

Hedrick PW. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol*. **22**:4606–4618.

Helfrich P, Rieb E, Abrami G, Lücking A, Mehler A. 2018. TreeAnnotator: Versatile visual annotation of hierarchical text relations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) [cited 2022 Jun 21]. Available from: https://aclanthology.org/L18-1308.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the quaternary. *Philos Trans R Soc Lond B Biol Sci* **359**: 183–195.

Hill AV. 1998. The immunogenetics of human infectious diseases. *Ann Rev Immunol*. **16**:593–617.

Hopwood AT, Hollyfield JP. 1954. *An annotated bibliography of the fossil mammals of Africa (1742-1950)*. London: British Museum (Nat. Hist.).

Jori F, Vial L, Penrith ML, Pérez-Sánchez R, Etter E, Albina E, Michaud V, Roger F. 2013. Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. *Virus Res*. **173**: 212–227.

Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. *Nat Rev Genet*. **15**: 379–393.

Kingdon J. 2013. Mammalian evolution in Africa. In: Kingdon J, Happold DCD, Butynski TM, Hoffman M, Happold M, Kalina J, editors. *Mammals of Africa. Introductory chapters and Afrotheria*. Vol. 1. London: Bloomsbury Publishing. p. 75–100.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. *BMC Bioinform*. **15**:356.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinform*. **14**:289.

Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A*. **99**:803–808.

Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. **49**:W293–W296.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. **26**:589–595.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature*. **475**:493–496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**:2078–2079.

Libri V, Schulte D, van Stijn A, Ragimbeau J, Rogge L, Pellegrini S. 2008. Jakmip1 is expressed upon T cell differentiation and has an inhibitory function in cytotoxic T lymphocytes. *J Immunol* **181**:5847–5856.

Lin CS, Sun YL, Liu CY, Yang PC, Chang LC, Cheng IC, Mao SJ, Huang MC. 1999. Complete nucleotide sequence of pig (*Sus scrofa*) mitochondrial genome and dating evolutionary divergence within Artiodactyla. *Gene*. **236**:107–114.

Liu L, Bosse M, Megens H-J, Frantz LAF, Lee Y-L, Irving-Pease EK, Narayan G, Groenen MAM, Madsen O. 2019. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat Commun*. **10**:1992.

Lorenzen ED, Arctander P, Siegismund HR. 2006. Regional genetic structuring and evolutionary history of the impala Aepyceros melampus. *J Hered*. **97**:119–132.

Lorenzen ED, Heller R, Siegismund HR. 2012. Comparative phylogeography of African savannah ungulates. *Mol. Ecol*. **21**:3656–3670.

Maestri A, Sortica VA, Tovo-Rodrigues L, Santos MC, Barbagelata L, Moraes MR, Alencar de Mello W, Gusmão L, Sousa RCM, Emanuel Batista Dos Santos S. 2015. Siaα2-3Galβ1- receptor genetic variants are associated with influenza A(H1N1)pdm09 severity. *PLoS One*. **10**:e0139681.

Martin SH, Davey JW, Jiggins CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol*. **32**: 244–257.

McBride KE, Cheemarla NR, Guerrero-Plata MA. 2018. Role of mucin 19 in the respiratory tract. *J Immunol*. **200**:60.8.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al*. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**:1297–1303.

Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*. **210**: 719–731.

Meisner J, Albrechtsen A. 2019. Testing for Hardy–Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Mol Ecol Resour*. **19**: 1144–1152.

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, *et al*. 2021. Sustainable data analysis with snakemake. *F1000Res*. **10**:33.

Moritz C. 1994. Defining "Evolutionarily Significant Units" for conservation. *Trends Ecol Evol*. **9**:373–375.

Muwanika VB, Nyakaana S, Siegismund HR, Arctander P. 2003. Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*. **91**: 361–372.

Nersting LG, Arctander P. 2001. Phylogeography and conservation of impala and greater kudu. *Mol Ecol*. **10**:711–719.

Netherton CL, Simpson J, Haller O, Wileman TE, Takamatsu H-H, Monaghan P, Taylor G. 2009. Inhibition of a large double-stranded DNA virus by MxA protein. *J Virol*. **83**:2310–2320.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS One*. **7**(7):e37558.

Nimmerjahn F, Ravetch JV. 2006. Fcgamma receptors: Old friends and new family members. *Immunity*. **24**:19–28.

Ní Leathlobhair M, Perri AR, Irving-Pease EK, Witt KE, Linderholm A, Haile J, Lebrasseur O, Ameen C, Blick J, Boyko AR, et al. 2018. The evolutionary history of dogs in the Americas. *Science.* **361**:81–85.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature.* **499**(7456):74–78.

Pacifici M, Santini L, Di Marco M, Baisero D, Francucci L, Marasini GG, Visconti P, Rondinini C. 2013. Generation length for mammals. *Nat Conserv.* **5**:89.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics.* **192**:1065–1093.

Pečnerová P, Garcia-Erill G, Liu X, Nursyifa C, Waples RK, Santander CG, Quinn L, Frandsen P, Meisner J, Stæger FF, et al. 2021. High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Curr Biol.* **31**:1862–1871.e5.

Pedersen C-ET, Albrechtsen A, Etter PD, Johnson EA, Orlando L, Chikhi L, Siegismund HR, Heller R. 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nat Ecol Evol.* **2**:491–498.

Peter BM, Slatkin M. 2013. Detecting range expansions from genetic data. *Evolution.* **67**:3274–3289.

Petkova D, Novembre J, Stephens M. 2016. Visualizing spatial population structure with estimated effective migration surfaces. *Nat Genet.* **48**:94–100.

Pickford M. 2006. Synopsis of the biochronology of African neogene and quaternary suiformes. *Trans R Soc S Afr.* **61**:51–62.

Pickford M. 2012. Ancestors of Broom's pigs. *Trans R Soc S Afr.* **67**:17–35.

Pickford M. 2013a. The diversity, age, biogeographic and phylogenetic relationships of Plio-Pleistocene suids from Kromdraai, South Africa. *Ann Ditsong Natl Museum Nat History.* **3**:11–32.

Pickford M. 2013b. Locomotion, diet, body weight, origin and geochronology of *Metridiochoerus andrewsi* from the Gondolin karst deposits, Gauteng, South Africa. *Ann Ditsong Natl Museum Nat History* **3**:33–47.

Pickford M, Gommery D. 2016. Fossil Suidae (Artiodactyla, Mammalia) from Aves Cave I and nearby sites in Bolt's Farm Palaeokarst System, South Africa. *Estudios Geologicos (Madrid).* **72**:059.

Pickford M, Gommery D. 2020. Fossil suids from Bolt's Farm Palaeokarst System, South Africa: Implications for the taxonomy of *Potamochoeroides* and *Notochoerus* and for biochronology. *Estudios Geologicos (Madrid).* **76**:127.

Pickrell J, Pritchard J. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**:e1002967.

Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K. 2020. GenMap: Ultra-fast computation of genome mappability. *Bioinformatics.* **36**:3687–3692.

Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, et al. 2016. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell.* **167**:643–656.e17.

Racimo F, Marnetto D, Huerta-Sánchez E. 2017. Signatures of archaic adaptive introgression in present-day human populations. *Mol Biol Evol.* **34**:296–317.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing [cited 2022 Jun 21]. Available from: https://www.R-project.org/.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol.* **67**:901–904.

Randi E, D'Huart J-P, Lucchini V, Aman R. 2002. Evidence of two genetically deeply divergent species of warthog, *Phacochoerus africanus* and *P. aethiopicus* (Artiodactyla: Suiformes) in East Africa. *Mamm Biol.* **67**:91–96.

Ricciotti E, FitzGerald GA. 2011. Prostaglandins and inflammation. *Arterioscler Thromb Vasc Biol.* **31**:986–1000.

Sander WJ, O'Neill HG, Pohl CH. 2017. Prostaglandin E2 as a modulator of viral infections. *Front Physiol.* **8**:89.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature.* **507**:354–357.

Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, Munters AR, Vicente M, Steyn M, Soodyall H, et al. 2017. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science.* **358**:652–655.

Sjödin P, McKenna J, Jakobsson M. 2021. Estimating divergence times from DNA sequences. *Genetics.* **217**(4):iyab008.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics.* **195**:693–702.

Souron A. 2016. On specimens of extant warthogs (*Phacochoerus*) from the Horn of Africa with unusual basicranial morphology: rare variants of *Ph. africanus* or hybrids between *Ph. africanus* and *Ph. aethiopicus*? *Suiform Soundings* **15**:86–92.

Souron A. 2017. Diet and ecology of extant and fossil wild pigs. In: Melletti MME, editor. *Ecology, conservation and management of wild pigs and Peccaries.* Cambridge, UK: Cambridge University Press. p. 29–38.

Taylor HC. 1977. Aspects of the ecology of the Cape of Good Hope Nature Reserve in relation to fire and conservation. US Forest Service, Washington Office, General Technical Report. p. 483–487.

Van Rossum G, Drake FL. 2009. *Python 3 reference manual.* Scotts Valley, CA: CreateSpace.

Vercammen P, Mason DR. 1993. The warthogs (*Phacochoerus africanus* and *P. aethiopicus*). In Oliver WLR, editor. Pigs, peccaries and hippos: status survey and action plan. Gland, Switzerland: IUCN. p. 75–84.

Verhelst J, Hulpiau P, Saelens X. 2013. Mx proteins: Antiviral gatekeepers that restrain the uninvited. *Microbiol Mol Biol Rev.* **77**:551–566.

Wang S, Zhang J, Zhang Y, Yang J, Wang L, Qi Y, Han X, Zhou X, Miao F, Chen T, et al. 2020. Cytokine storm in domestic pigs induced by infection of virulent African Swine Fever Virus. *Front Vet Sci* **7**:601641.

Waples RK, Albrechtsen A, Moltke I. 2019. Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Mol Ecol.* **28**:35–48.

White TD, Harris JM. 1977. Suid evolution and correlation of African hominid localities. *Science.* **198**:13–21.

Wu G-S, Yao Y-G, Qu K-X, Ding Z-L, Li H, Palanichamy MG, Duan Z-Y, Li N, Chen Y-S, Zhang Y-P. 2007. Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biol.* **8**:R245.

Yang Z, Donoghue PCJ. 2016. Dating species divergences using rocks and clocks. *Philos Trans R Soc Lond B Biol Sci.* **371**:20150126.

Zhang M, Yang Q, Ai H, Huang L. 2022. Revisiting the evolutionary history of pigs via de novo mutation rate estimation in a three-generation pedigree. *Genomics Proteomics Bioinf.* doi:10.1016/j.gpb.2022.02.001.

Zhou J, Chen J, Zhang X-M, Gao Z-C, Liu C-C, Zhang Y-N, Hou J-X, Li Z-Y, Kan L, Li W-L, et al. 2018. Porcine Mx1 protein inhibits classical swine fever virus replication by targeting nonstructural protein NS5B. *J Virol.* **92**:e02147–17.

Zhu JJ, Ramanathan P, Bishop EA, O'Donnell V, Gladue DP, Borca MV. 2019. Mechanisms of African swine fever virus pathogenesis and immune evasion inferred from gene expression changes in infected swine macrophages. *PLoS One.* **14**:e0223955.