Title: Genomic and morphometric evidence for Austronesian-mediated pig translocation in the Pacific

Author list: David W. G. Stanton[1,2] †, Aurelie Manin[3,4] †, Allowen Evin[5] †, Kristina Tabbada[3] †, Anna Linderholm[6,7,3] †, Rosie Drinkwater[1,8], Olaf Thalmann[8], Said I. Ng'ang'a[1,8], Noel Amano[9], Atholl Anderson[10], Ross Barnett[11], Patrick Barrière[12], Stuart Bedford[13,14], Peter Bellwood[13], Adam Brumm[15,16], Trung Cao Tien[17,18], Geoffrey Clark[10], Richard Crooijmans[19], Thomas Cucchi[20], Michelle Eusebio[21], Linus Girdland-Flink[22], Peter Galbusera[23], Martien Groenen[19], Budianto Hakim[24], Stuart Hawkins[13], Holly Heiniger[25,26], Kristofer M. Helgen[27], Michael J. Herrera[21], Mary Jane Louise A. Bolunia[28], Eusebio Z. Dizon[28] ‡, Hsiao-chun Hung[10], Karen Greig[29], Terry Hunt[30], Andrew C. Kitchener[31,32], Carol Lee[33,34], Alastair A. Macdonald[35], Hendrik-Jan Megens[19], Erik Meijaard[36,37], Kieren J. Mitchell[25,26], Christopher Moran[33], Karen Mudar[38], Karma Nidup[33], Marc Oxenham[13,39], Rinzin Pem[40], Philip J. Piper[13], Lachie Scarsbrook[3,8], Kyle Schachtschneider[41], Lawrence Schook[42], Pradeepa Silva[43], Matthew Spriggs[13,44], Samuel T. Turvey[45], Una S. Viðarsdóttir[46], Murray P. Cox[47,48], Tim Denham[13,] Jaime Gongora[33], Keith Dobney[22,49], Greger Larson[3] *, Laurent Frantz[1,8] *

1 School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK
2 School of Biosciences, Cardiff University, Cardiff, UK
3 Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, UK
4 PACEA, Université de Bordeaux-CNRS, Bordeaux, France
5 Institute of Evolutionary Science-Montpellier (ISEM), University of Montpellier, CNRS, IRD, EPHE, Montpellier, France
6 Centre for Palaeogenetics, Stockholm University, Stockholm, Sweden
7 Department of Geological Sciences, Stockholm University, Stockholm, Sweden
8 Palaeogenomics Group, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany
9 Max Planck Institute of Geoanthropology, Jena, Germany
10 Archaeology and Natural History, School of Culture, History and Language, Australian National University, Canberra, Australia
11 Globe Institute, University of Copenhagen, Copenhagen, Denmark
12 Agence néo-Calédonienne de la Biodiversité, Pôle Menaces, Presqu'ile de Foué, Koné, New Caledonia, France
13 School of Archaeology and Anthropology, College of Arts and Social Sciences, Australian National University, Canberra, ACT, Australia
14 Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
15 Australian Research Centre for Human Evolution (ARCHE), Griffith University, Brisbane, Australia
16 School of Environment and Science, Griffith University, Brisbane, Australia
17 School of Chemistry, Biology and Environment, Vinh university, Vinh, Vietnam
18 Institute for Ecology and Conservation of Nature, Vinh University, Vinh, Vietnam
19 Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands
20 BioArch, UMR 7209, Muséum National d'Histoire Naturelle/CNRS, Paris, France
21 School of Archaeology, University of the Philippines Diliman, Quezon City, The Philippines
22 Department of Archaeology, School of Geosciences, University of Aberdeen, Aberdeen, UK
23 Antwerp Zoo Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium
24 Pusat Riset Arkeologi Prasejarah dan Sejarah, Badan Riset dan Inovasi Nasional (BRIN), Makassar, Indonesia
25 School of Biological Sciences, University of Adelaide, Adelaide, Australia
26 Manaaki Whenua – Landcare Research, Lincoln, Canterbury, New Zealand
27 Australian Museum Research Institute, Sydney, Australia
28 Archaeology Division, National Museum of the Philippines, Manila, Philippines
29 Archaeology Programme, School of Social Sciences, University of Otago, Dunedin, New Zealand
30 School of Anthropology, University of Arizona, Tucson, USA
31 Department of Natural Sciences, National Museums Scotland,Chambers Street, Edinburgh, UK
32 School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh, UK
33 Sydney School of Veterinary Science, Faculty of Science, University of Sydney, Australia
34 Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, New South Wales, Sydney, Australia
35 Royal (Dick) School of Veterinary Studies & The Roslin Institute, University of Edinburgh, Edinburgh, UK
36 Borneo Futures, Bandar Seri Begawan, Brunei Darussalam, Brunei
37 Durrell Institute of Conservation and Ecology, University of Kent, Canterbury, UK
38 National Park Service, Washington DC, USA
39 School of Geosciences, University of Aberdeen, Aberdeen, UK
40 Department of Livestock, Ministry of Agriculture and Livestock, Bhutan
41 Department of Radiology, University of Illinois at Chicago, Chicago, USA
42 Department of Animal Sciences, University of Illinois, Urbana-Champaign, USA
43 Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka
44 Vanuatu Cultural Centre, Port Vila, Vanuatu
45 Institute of Zoology, Zoological Society of London, London, UK
46 Biomedical Center, University of Iceland, Reykjavik, Iceland
47 Department of Statistics, University of Auckland, Auckland, New Zealand
48 College of Sciences, Massey University, Palmerston North, New Zealand
49 Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK
* Lead contact
† These authors contributed equally
‡ Deceased

Abstract: Several millennia of human mediated translocation of non-native pig species (genus *Sus*) to the islands of Wallacea and Oceania has significantly altered local ecosystems. To investigate the timing and trajectory of these introductions, we conducted both genomic analyses of 576 pig nuclear genomes and a geometric morphometric analysis of 708 modern and ancient dental remains. Our analyses demonstrate that free-living and domestic pigs in Wallacea and Oceania possess diverse ancestries resulting from the introduction of multiple, sequential pig populations followed by gene flow. Despite the variability in their genomic ancestry, these pigs all possess a distinct tooth morphology, and a genetic link to the Chinese domestic pig populations that accompanied the dispersal of Austronesian language speakers ~4,000-3,000 years ago via Taiwan and the Philippines.

Introduction

People have transported animal species both involuntarily (e.g., commensal species such as rats and mice), and deliberately (e.g., as livestock or part of game management practices) beyond their native ranges for millennia. These introductions have often substantially altered local ecosystems, particularly on islands (*1*). The translocation of vertebrates eastwards across the Wallace Line (Fig. 1A) into Wallacea, a biogeographic region between the Asian and Australasian biotas, has had dramatic environmental impacts (*2*). For example, while the natural range of the genus *Sus* (pigs) is primarily located west of the Wallace Line, and limited to the westernmost part of Wallacea (i.e., Sulawesi and other islands between the Asian and Australasian continental regions), pigs are now widespread across archipelagoes in Wallacea, and Near and Remote Oceania (Fig. 1A). The geographic origin and ancestry of these pig populations across Wallacea and Oceania, however, remain unknown.

Previous genomic research has shown that pigs within the genus *Sus* can be broadly classified into three groups: Western Eurasian wild boar (*S. scrofa*, the ancestor of Western domestic pigs), Eastern Eurasian wild boar (*S. scrofa*, the ancestor of Eastern domestic pigs), and Island Southeast Asian (ISEA) endemic species found in Borneo and Sumatra (*S. barbatus*), Java (*S. verrucosus*), Sulawesi (*S. celebensis*), and the Philippines (*S. cebifrons*) (Fig. 1A; (*3–6*)). The human-mediated translocation of pigs east of the Wallace Line most likely began with *S. celebensis* in pre-Neolithic times >4,000 years ago, prior to the dispersal of Austronesian language speaking peoples and domestic animals into the region (*7*). The recent discovery on Sulawesi of the world's oldest figurative artwork (>50,000 years) that depicts the local endemic warty pig species (*S. celebensis*) demonstrates the deep-rooted relationship between humans and pigs in the region (*8*, *9*). Mitochondrial DNA and zooarchaeological analyses have identified *S. celebensis* remains from pre-Neolithic layers (purportedly older than 4,000 years) in Liang Bua Cave, Flores (*10–12*), though these have not yet been directly dated. If confirmed by radiocarbon dating, the presence of this species in pre-Neolithic deposits, ~300 km from its natural range in Sulawesi across the Flores Sea, would suggest these pigs were intentionally translocated by people.

Mitochondrial analyses indicated that the Eurasian pig (*S. scrofa*) also occurred in later deposits at Liang Bua Cave (younger than 4,000 years ago), suggesting *S. scrofa* was introduced to Flores subsequent to the introduction of *S. celebensis* (*10–12*). In fact, analyses of pig mitochondrial data from Wallacea and Oceania showed that introduction of *S. scrofa* was considerably more widespread than *S. celebensis*, since populations of *S. scrofa* are now free-living throughout these regions (Fig. 1; (*10*)). These *S. scrofa* populations were likely first introduced east of the Wallace Line 4,000-3,000 years ago during the expansion of Austronesian speakers from mainland East Asia and Taiwan, through the Philippines and Northern Wallacea, into Melanesia and Polynesia (*13*). This is supported by the remains of pigs (~2,900 years before present [BP]) excavated in Vanuatu from Lapita archaeological contexts associated with the Austronesian expansion (*14–16*).

Pigs from Oceania and Wallacea predominantly possess both a particular tooth shape and a distinct mitochondrial haplogroup known as the "Pacific Clade" (*10*, *17*). This mtDNA signature has also been identified in free-living *S. scrofa* individuals from Java, Sumatra, and mainland Southeast Asia, including Laos, Vietnam, and Yunnan Province in southwestern China. This distribution has been interpreted to mean that pigs introduced to Wallacea likely originated from peninsular Southeast Asia (*10*, *17*). In contrast, modern domestic pigs found along the hypothesized route of the Austronesian expansion in Southeast China, Taiwan, and the Philippines possess distinct Lanyu (known as the "Lanyu Clade") and East Asian mitochondrial signatures (*10*, *17*). These data suggest that although Austronesian speakers may have travelled with pigs that possessed the Lanyu Clade signature as they arrived in the Philippines, they only acquired pigs with the Pacific Clade mitochondrial haplotype, as well as a variety of horticultural and arboreal crops, including banana, sugarcane and yam (*18*), after leaving the Philippines.

West of the Wallace Line, however, the Pacific Clade mitochondrial haplogroup has only been identified in free-living pigs (*10*, *17*). The absence of this haplogroup in modern domestic pigs thus challenges the idea that it represents a marker of domestic pigs introduced from mainland Asia. Furthermore, recent nuclear genome analyses have shown that, although wild *S. scrofa* populations are primarily found in mainland Eurasia, some *S. scrofa* populations naturally occur in Western Indonesia (Sumatra, and Java; Fig. 1A; (*3*, *4*)). While these populations west of the Wallace Line mostly possess haplogroups typically associated with ISEA-endemic *Sus* (due to gene flow; (*3*, *4*)), some individuals possess the Pacific Clade haplogroup (*10*, *17*). It thus remains unclear whether free-living pigs found east of the Wallace Line represent *S. scrofa* populations that dispersed naturally (analogous to *S. celebensis*), or were translocated eastward by people.

The successful natural dispersal of the ancestor of another ISEA-endemic *Sus, S. celebensis*, from Borneo to Sulawesi (*4*), and its subsequent human-mediated translocation to the Lesser Sunda Islands suggests that a combination of natural and human-mediated dispersal of wild *S. scrofa* across the Wallace Line and onward toward Oceania is plausible. In this scenario, the pigs discovered in Lapita archaeological contexts may have been the descendants of domestic pigs introduced by Austronesian-speaking groups from mainland Asia, and that they acquired mitochondrial haplotypes belonging to the Pacific Clade through gene flow with other pig populations that were already present in Wallacea.

To establish the geographic origin of the pig populations in Wallacea, Melanesia, Micronesia and Polynesia, we sequenced 117 *Sus* nuclear genomes and generated geometric morphometric data (from the third lower molar) for 401 modern and 313 archaeological specimens. The genomic data generated here included 54 modern genomes, 60 historical genomes (1871-1993) and three ancient genomes from the Philippines (~500 years BP), Palau (Micronesia; ~1400 years BP), and the Lapita archaeological site of Teouma on Efate Island, Vanuatu (~2,900 years BP). We analyzed these data alongside publicly available data that included 400 nuclear genomes (Fig. 1A) and ~700 mtDNA control region sequences.

Results and Discussion

The origin of the Pacific Clade mitochondrial haplogroup and associated morphometric signature

We assessed the spatiotemporal frequency of both the Pacific Clade and Lanyu Clade across ISEA and the Pacific, based on a data set consisting of 585 mtDNA whole genome (>10x coverage; Fig. S3), 703 control region fragment sequences including historical and ancient specimens (*10*, *17*), and recently published datasets of 356 modern individuals from the Philippines (*19*, *20*). We identified 58 individuals that possessed mitochondrial sequences belonging to the Lanyu haplogroup, all of which were found in the Philippines (n=52, of which n=7 from Palawan) and Taiwan (n=6). Of the 90 individuals that possessed a Pacific Clade signature (Fig. 1B), the majority (n=55) were located east of the Wallace Line in Wallacea, New Guinea, Melanesia, Polynesia, and in a single domestic pig from the Philippines (out of 356). The majority (~77%) of pigs east of Wallace line possessed the Pacific Clade, while ~22% (16/72) possessed haplotypes belonging to

the East Asian haplogroup and one individual from New Caledonia possessed a haplotype belonging to the European clade. The remaining individuals with a Pacific Clade haplogroup were found across South and Peninsular Southeast Asia (Fig. 1B).

These results suggest that the distinctive tooth shape (Pacific shape) previously associated with pigs carrying a Pacific Clade mtDNA signature may also be present west of the Wallace Line. To address this, we conducted a morphometric analysis of pig teeth (third lower molar) based on 307 archaeological and 401 modern specimens. We used 12 landmarks and 87 sliding semilandmarks to capture tooth shape variation. The Pacific shape was identified in pigs east of the Wallace Line including modern and ancient individuals from Papua New Guinea (PNG), Vanuatu and the Marquesas Islands (Fig. 1C). This tooth shape is distinct from the shape observed in archaeological pigs from mainland East Asia (correct cross-validation = 97% (CI: 92.3-100%); Data S3).

None of the 36 archaeological individuals from the Philippines possessed the Pacific shape, while four archaeological out of 35 individuals in total (both modern and archaeological) from Taiwan were identified as having this shape with high probability (>92%; Data S3). Nine out of 14 individuals from Sarawak (Borneo) were also found to possess the Pacific shape. These results demonstrate that multiple pig populations found west of the Wallace line in Taiwan, the Philippines, Borneo, and mainland Asia carry the distinctive Pacific mtDNA and tooth shape signatures. This geographic distribution, and the possibility of morphological convergence, complicates the use of these markers for determining the origins of pigs found east of the Wallace Line.

The ancestry of pigs east of the Wallace Line

To further investigate the origin of pigs east of the Wallace Line, we analyzed 576 nuclear genomes sequenced to >0.1x coverage (Data S1). We first conducted a Principal Component Analysis (PCA) that corroborated the existence of three major ancestry clines (Fig. 1D) observed in previous publications (*4*, *21*, *22*). Principal Component 1 (PC1) differentiated the reciprocally monophyletic *S. scrofa* (wild and domestic) and wild ISEA-endemic *Sus* species: *S. celebensis*, *S. cebifrons*, *S. philippensis*, *S. barbatus* and *S. verrucosus* (Fig. 1D). PC2 differentiated European (wild and domestic) and Asian *S. scrofa* populations (wild and domestic; Fig. 1D). *S. scrofa* pigs from ISEA (including Wallacea), Micronesia, Melanesia, and Polynesia (represented by red triangles in Fig. 1A & D) are broadly distributed across PC1 and PC2, indicating that they possess mixed ancestry that includes ISEA-endemic *Sus*, Asian *S. scrofa* and European *S. scrofa*. These results are consistent with the observation that the majority of free-living *S. scrofa* individuals west of the Wallace Line from Sumatra, Java, the Malay Peninsula and nearby islands, also possessed ISEA-endemic *Sus* mtDNA haplogroups (Fig. 1B; Fig. S3).

To quantify the proportion of ancestry derived from ISEA-endemic *Sus*, Asian *S. scrofa* and European *S. scrofa* proportions in individual genomes, we used a combination of analyses including ADMIXTURE and D-statistics. We first employed an unsupervised ADMIXTURE analysis (K=3) based on 50 wild *S. scrofa* individuals from Europe and 40 from East Asia (China and Korea), in addition to the 30 ISEA-endemic *Sus* genomes in our dataset. This facilitated the identification of representative genomes for each ancestry (i.e., those that possessed >99% from only one source; Fig. S6). These representatives served as sources in a supervised ADMIXTURE (K=3) analysis that included the 576 genomes with >0.1x coverage. We then decomposed ISEA-endemic *Sus* ancestry into one of four components (*S. celebensis*, *S. cebifrons/S. philippensis*, *S. barbatus*, and *S. verrucosus*) using D-statistics (see Supplementary Material).

These analyses confirmed the widespread of both ISEA-endemic *Sus* and Asian *S. scrofa* nuclear ancestry in free-living *S. scrofa* populations east and west of the Wallace Line (*4*). Many of these samples also possessed ISEA-endemic *Sus* mtDNA haplogroups, particularly in Wallacea, Sumatra, Java, and the Malay

Peninsula (Fig. 1B; Fig. 2). There was also a strong concordance between geographic location and ISEA-endemic *Sus* ancestry: 18 *S. scrofa* individuals from Java and Sumatra possessed between ~13-37% ancestry derived from either *S. verrucosus* (endemic to Java) or *S. barbatus* (endemic to Sumatra and Borneo), while eight admixed individuals east of the Wallace Line possessed between ~15-80% *S. celebensis* ancestry (endemic to Sulawesi, east of the Wallace Line; Fig. 2). We found little to no *S. cebifrons/S. philippensis* ancestry (endemic to the Philippines) in 14 modern domestic pigs (*S. scrofa*) from the Philippines (Fig. 2).

Most free-living *S. scrofa* from Sumatra, Java, and nearby islands possessed similar levels of *S. verrucosus* ancestry (13-18%). This finding supports previous studies indicating that wild *S. scrofa* from Sumatra interbred with *S. verrucosus* during the Pleistocene when low sea levels allowed for the connection of Sumatra, Borneo, and Java via the exposed Sunda Shelf, which facilitated migration and population exchange (*3*, *4*). Three wild pigs from Simeulue Island, northwest of Sumatra, which were suspected to have been part of a population translocated from Sulawesi based on their morphological classification as *S. celebensis* (*23*), possessed ~30% *S. barbatus* and ~70% Asian *S. scrofa* ancestry. These results suggest that gene flow can lead to the generation of hybrid populations that can then confuse morphological analyses, resulting in spurious species identifications.

Although all 12 individuals in Wallacea i.e. free-living pigs from Maluku, the Lesser Sundas and New Guinea are outside of the natural range of *Sus* species they all possessed some degree of Asian *S. scrofa* and/or *S. celebensis* ancestry, though the proportion of each varied widely between islands (Fig. 2). In the Maluku archipelago northeast of Sulawesi, *S. celebensis* ancestry in free-living pigs ranged from 80% on Halmahera to ~5% on Morotai, just a few kilometers north. Despite its high proportion of *S. celebensis* ancestry, this 1979 individual from Halmahera was classified and accessioned as *S. scrofa* at the Smithsonian Natural History Museum (*24*). Additionally, *S. celebensis* ancestry was absent in free-living pigs from the Central Maluku islands of Buru and Seram, despite previous surveys suggesting the presence of *S. celebensis* on Buru (*25*). This lack of *S. celebensis* ancestry on Buru is also puzzling considering the presence of babirusa (*Babyrousa babyrussa*) on the island, which are thought have originated from Sulawesi, where *S. celebensis* also occurs.

A similar pattern was observed in the Lesser Sunda region of Wallacea including the islands of Timor, where *S. celebensis* ancestry in free-living pigs ranged from 100% to 29%, and Flores, where it ranged from 16% to 5%. These results demonstrate that *S. celebensis* populations occur well beyond its native range on Sulawesi. While natural dispersal to some islands cannot be ruled out, the open water distance between Sulawesi and Flores suggests that seafaring humans deliberately translocated this species during or since the Late Pleistocene (*10*, *11*). Despite the significant variation in *S. celebensis* ancestry (0-80%) in individuals found east of the Wallace Line, the consistent presence of a *S. scrofa* Pacific Clade mtDNA haplogroup across all individuals, indicates a component of Asian *S. scrofa* ancestry (Fig. 1B).

We did not detect any ISEA-endemic *Sus* ancestry in domestic pigs from the Philippines, nor in free-living or domestic pigs from Melanesia or Polynesia. European *S. scrofa* ancestry was also absent from historical, free-living individuals from New Guinea dated to the early-to-mid-20th century, but ranged from 10% to 58% in modern domestic New Guinea populations. In modern New Caledonian free-living pigs, European ancestry ranged from ~43-57%, and a free-living pig from Hawaii, sampled in 1937, possessed 58% European ancestry. All recent domestic pigs from the Philippines possessed between 27-90% European ancestry. We also detected ~42% European ancestry in a free-living pig from Sumba, a Wallacean island south of Flores.

Patterns of European and Asian *S. scrofa* ancestry covariance across the genome based on DATES (*26*) indicated that European *S. scrofa* ancestry was introduced to these islands during the 19th and early 20th

centuries (Supplementary Material). In addition, ancient genomes from Vanuatu (~2,900 BP), the Philippines (~500 BP), and Palau (~700 BP) showed no evidence of European ancestry. These results demonstrate that European *S. scrofa* ancestry was introduced to these regions, likely via European domestic pigs during and since the colonial period, and has since risen to substantial proportions in both free-living and domestic pig populations across the region.

The geographic origin of Asian S. scrofa genomic ancestry east of the Wallace Line

Although multiple historical/modern free-living and domestic pigs east of the Wallace Line possess a diverse mix of ancestry, including *S. celebensis* and European *S. scrofa*, they all possess some degree of Asian *S. scrofa* ancestry, and often possess a *S. scrofa*-like mtDNA haplogroup (Fig. 1B). Assessing the geographic origin of this Asian *S. scrofa* ancestry is challenging, however, particularly in mixed genomes that also feature European *S. scrofa* or ISEA-endemic *Sus* ancestry. In addition, the Asian *S. scrofa* lineage is widespread and encompasses wild boar from North and South China, Korea, Japan, ISEA (Sumatra and Java), as well as ancient, historical, and modern pigs from the Philippines, Near and Remote Oceania, and East Asian domestic pigs.

To overcome these constraints, we performed ancestry deconvolution, a process by which specific ancestries in admixed individuals can be identified and isolated using local ancestry inference. We first phased all genomes with >8x coverage using SHAPEIT v4 (*27*), and imputed the genomes with >1x coverage using GLIMPSE (*28*). We then used GNOMIX (*29*) to identify ancestry blocks from Asian *S. scrofa*, European *S. scrofa*, and ISEA-endemic *Sus* ancestry, using wild, un-admixed individuals as reference populations (see Supplementary Material). Downsampling showed that GNOMIX local ancestry inference is highly accurate (94-99%) for genomes over 0.1x coverage (Fig. S5).

For each admixed genome, we filtered ancestry blocks assigned to European *S. scrofa* and ISEA-endemic *Sus* ancestry. We then computed f4 ratios to compare levels of non-Asian *S. scrofa* ancestry before and after filtering out both European *S. scrofa* and ISEA-endemic *Sus* ancestry blocks (Fig. 3A). European *S. scrofa* ancestry in admixed individuals from New Caledonia, Hawaii, Papua New Guinea and Taiwan (Lanyu pigs) decreased from ~25-50% to 0-5%, while European *S. scrofa* ancestry in domestic pigs from the Philippines decreased from ~30-70% to 0-15%. ISEA-endemic *Sus* (*S. celebensis*) ancestry in individuals from Wallacea, including Flores, Timor, and Halmahera, decreased from between ~40-80% to below 5%.

ISEA-endemic *Sus* (*S. verrucosus*) ancestry in wild pigs from west of the Wallace Line, in Java, Sumatra, and nearby islands, however, decreased from between 5-14% to ~5%. This slightly lower deconvolution success in these individuals can be explained by ancient admixture that took place during the Pleistocene (*3*, *4*). Because this gene flow took place earlier in time, the ISEA-endemic *Sus* ancestry across the genome now exists in blocks that are too small to identify. As a result of its more recent introduction to the region, the European *S. scrofa* ancestry is present in larger blocks (Fig. 3B), and is thus more easily identified. The wild pigs from west of the Wallace Line likely possess *S. scrofa* ancestry that is not well represented in our panels, which could make it harder to identify by GNOMIX (see Supplementary Material).

Principal component analysis using deconvoluted genomes indicated that populations east of the Wallace Line possessed a closer genetic affinity with mainland domestic populations—such as breeds from Southeast China (such as Luchuan, Wuzhishan, and Bamaxiang)—as well as with Lanyu pigs from Taiwan and domestic pigs from the Philippines, which aligns with the hypothesized route of the dispersal of Austronesian language speakers (*13*).

We then assessed the genetic similarities between the Asian-derived ancestry of *S. scrofa* individuals east of the Wallace Line and a 0.7x coverage genome from a ~2,900-year-old pig specimen (POU_Vanuatu_3 in

Fig. 1A) excavated at a Lapita archaeological site (associated with Austronesian speakers) in Vanuatu (*14*). The relatively low coverage of this ancient individual (0.7x) makes direct comparison with deconvoluted genomes challenging, since the deconvolution process itself (that removes European *S. scrofa* or ISEA-endemic *Sus* ancestry) further reduces the effective coverage.

To circumvent this issue, we first computed outgroup-f3 statistics comparing POU_Vanuatu_3 with modern samples that only possessed Asian *S. scrofa* ancestry. The objective of this step was to identify a proxy individual with >1x coverage suitable for subsequent comparisons against all samples, including the deconvoluted individuals. We found POU_Vanuatu_3 shared most drift with other Wallacean individuals and domestic pigs from Southeast China (such as Luchuan, Wuzhishan, Bamaxiang), and was closest to ISW_PNG_18, a ~100-year-old free-living pig from PNG (Fig. S11). Morphometric analyses further demonstrated that both ancient Vanuatu pigs from similar contexts and ISW_PNG_18 possessed the distinctive Pacific Shape tooth morphology (Fig. 1C; Data S1).

We therefore chose ISW_PNG_18 as a high-coverage proxy for the ancient Vanuatu specimen's ancestry. Outgroup-f3 analyses performed on the deconvoluted genome dataset, which included all Asian (mainland and ISEA) and Oceanian pigs, demonstrated that all *S. scrofa* individuals from Wallacea, Micronesia, Melanesia, and Polynesia share significantly stronger genetic affinities with the ISW_PNG_18 proxy compared to ISEA *S. scrofa* from west of the Wallace Line (Fig. 4). ISW_PNG_18 also shared more drift with ISU_UchuPalau_1, a 672-year-old Micronesian pig from Palau and ISU_Philippines_1, a 500-year-old individual from the Batanes Islands compared to individuals west and north of the Wallace line. This proxy also shared more drift with domestic pigs from Taiwan (Lanyu pigs), and South China (e.g., Bamaxiang, Luchuan and Wuzhishan pigs from Hainan Island and adjacent mainland Chinese provinces; underlined in Fig. 4) than with domestic pigs from Northern China and Vietnam, or ISEA populations west of the Wallace Line and wild Chinese *S. scrofa* (Fig. 4).

These results indicate that the domestic and free-living pigs found today beyond the natural range of *S. scrofa* in Wallacea and the Pacific, are closer genetically to ancient pigs from Vanuatu (from early Lapita archaeological contexts), and the Philippines, and modern domestic pigs in Taiwan, than to modern and historical individuals from peninsular Southeast Asia and western Indonesia. This indicates that pigs in the Pacific are, at least partly, descendants of domestic pigs introduced to the region via the Philippines, most probably by Austronesian-speaking groups who possessed East Asian human ancestry (*30*).

This demonstration of shared ancestry undermines the previously mitochondrially-based suggestion (*10*, *17*) that Polynesians pigs were not part of the Austronesian expansion since they were thought to have originated in Peninsular East Asia and dispersed through ISEA. The spatiotemporal distributions of the Pacific and Lanyu mitochondrial clades still show a substantial turnover in the mitochondrial haplotypes associated with Austronesian and Polynesian pigs. It is possible that both haplotypes were present (but as yet undetected) during the early phase of the Austronesian expansion, but due to recurrent bottlenecks associated with their dispersal, the Lanyu haplotypes vanished while the Pacific Clade haplotype became fixed.

Conditional heterozygosity analysis reveals that free-living and domestic pig populations east of the Wallace Line, and in Taiwan (i.e. Lanyu pigs) exhibit significantly lower heterozygosity compared to their mainland domestic ancestors (Fig. S14; $p<0.001$; Supplementary Material), suggesting that they likely experienced a bottleneck associated with founder events during their dispersal. Sequential bottlenecking could also explain the fixation of the rarer Pacific Clade haplotype. Alternatively, it is possible that the Pacific Clade haplotype was initially only present in native *Sus scrofa* populations in western ISEA. It then dispersed eastwards in populations that then hybridised with the Austronesian East Asian pigs whereupon the Pacific Clade

haplotype replaced the Lanyu signature. Additional ancient nuclear and mitochondrial DNA from ancient pigs across the region and from East Asia is necessary to determine which of these explanations is correct.

Multiple waves of dispersal and introduction in the Lesser Sunda Islands

It is noteworthy that a 1929 free-living pig from Sumbawa (ISW_Sumbawa_1), a few hundred kilometres east of the Wallace Line, possesses a mixture of Asian *S. scrofa* and *S. verrucosus* ancestry (endemic to Java), similar to pigs in Bali. Pigs from Flores and Timor (Fig. 2), however, possessed a mixture of Asian *S. scrofa* and *S. celebensis* ancestry (endemic to Sulawesi). Admixture modeling further suggests that at least part of the *S. scrofa* component in the Sumbawa pig is derived from a sister-lineage to Asian and European *S. scrofa*, similar to that observed in ISEA populations west of the Wallace Line (Fig. S9). These results suggest that Sumbawa pigs may possess ancestry from the western Lesser Sunda Islands, suggesting a possible natural and/or human-mediated translocation of pigs across the Wallace Line into the Lesser Sunda Islands. This ancestry pattern in the Lesser Sunda Islands illustrates the complex evolutionary history of pigs east of the Wallace Line.

Combined, our results suggest multiple dispersal processes in the region that possibly started with a natural and/or human-mediated translocation of ISEA pig populations that possessed *S. scrofa* and *S. verrucosus* ancestry from Bali moving east across the Wallace Line. This was likely followed by the introduction of *S. celebensis* from Sulawesi to Lesser Sunda Islands such as Flores and Timor, potentially by hunter-gatherer populations. Subsequently, domestic *S. scrofa* pigs derived from populations in mainland Asia were introduced into the region from northern Wallacea (following their introduction from southern China and the Philippines) sometime after 3,500-3,000 years ago. Many of these introduced pigs later became feral on islands across the region. Finally, pigs with European ancestry were introduced broadly across the region, including into the Lesser Sundas to the island of Sumba during the colonial period.

Conclusions

We found that most free-living and domestic pigs across a vast geographical region from the Philippines to Hawaii possess a degree of ancestry linked to domestic pigs (*S. scrofa*) likely introduced during the spread of Austronesian-speaking groups from Southeast China and Taiwan via the Philippines ~4,000 years ago. Interestingly, pigs in Near and Remote Oceania lack any detectable ancestry from wild endemic *Sus* species present along the Austronesian dispersal route in the Philippines, Sulawesi and other islands. This pattern suggests that the first wave of introduced domestic pigs did not interbreed with local, indigenous, *Sus* populations. Only their descendants left behind in Wallacea, and which subsequently became feral, had the opportunity to hybridize with indigenous wild species such as *S. celebensis* (Fig. 1A; Fig. 2).

The lack of gene flow between local *Sus* species and the first domestic pigs introduced to Wallacea is consistent with the pattern observed in the human populations who likely transported these pigs, and who did not mix with local groups during the initial phases of their dispersal to Melanesia (*30*). The continuing movement of Austronesian speakers eastwards across the islands of Oceania then resulted in the first establishment of domestic and feral pig (*S. scrofa*) populations across the region, where no pig species were previously present.

Their peculiar tooth morphology, and initial genetic isolation from local wild species they encountered in ISEA and Wallacea, may reflect the fact that these introduced *S. scrofa* pigs possessed specific domestic traits that facilitated their transport and management by Austronesian-speaking groups. Transporting these animals between islands resulted in a unique evolutionary history characterized by serial founder effects, gene flow from divergent lineages, and likely selection for specific traits that facilitated the establishment of feral populations. Future research utilizing high-coverage genomes from modern, historical, and archaeological samples, combined with functional assays and detailed morphometric analyses, will be crucial

to elucidate how these domestic pigs, introduced from mainland Asia, became so successful in establishing feral populations across ISEA and the Pacific.

The Materials and Methods are available in the Supplementary Materials.

References and Notes

1. D. R. Spatz, K. M. Zilliacus, N. D. Holmes, S. H. M. Butchart, P. Genovesi, G. Ceballos, B. R. Tershy, D. A. Croll, Globally threatened vertebrates on islands with invasive species. *Sci Adv* **3**, e1603080 (2017).

2. M. J. Struebig, S. G. Aninta, M. Beger, A. Bani, H. Barus, S. Brace, Z. G. Davies, M. D. Brauwer, K. Diele, C. Djakiman, R. Djamaluddin, R. Drinkwater, A. Dumbrell, D. Evans, M. Fusi, L. Herrera-Alsina, D. T. Iskandar, J. Jompa, B. Juliandi, L. T. Lancaster, G. Limmon, Lindawati, M. G. Y. Lo, P. Lupiyaningdyah, M. McCannon, E. Meijaard, S. L. Mitchell, S. Mumbunan, D. O'Connell, O. G. Osborne, A. S. T. Papadopulos, J. S. Rahajoe, Rosaria, S. J. Rossiter, Rugayah, H. Rustiami, U. Salzmann, Sheherazade, I. M. Sudiana, E. Sukara, J. S. Tasirin, A. Tjoa, J. M. J. Travis, L. Trethowan, A. Trianto, T. Utteridge, M. Voigt, N. Winarni, Z. Zakaria, D. P. Edwards, L. Frantz, J. Supriatna, Safeguarding Imperiled Biodiversity and Evolutionary Processes in the Wallacea Center of Endemism. *Bioscience* **72**, 1118–1130 (2022).

3. L. A. F. Frantz, O. Madsen, H.-J. Megens, M. A. M. Groenen, K. Lohse, Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian Sus species during the Plio-Pleistocene climatic fluctuations. *Mol. Ecol.* **23**, 5566–5574 (2014).

4. L. A. F. Frantz, J. G. Schraiber, O. Madsen, H.-J. Megens, M. Bosse, Y. Paudel, G. Semiadi, E. Meijaard, N. Li, R. P. M. A. Crooijmans, A. L. Archibald, M. Slatkin, L. B. Schook, G. Larson, M. A. M. Groenen, Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol.* **14**, R107 (2013).

5. L. Frantz, E. Meijaard, J. Gongora, J. Haile, M. A. M. Groenen, G. Larson, The Evolution of Suidae. *Annu Rev Anim Biosci* **4**, 61–85 (2016).

6. Y. Paudel, O. Madsen, H.-J. Megens, L. A. F. Frantz, M. Bosse, R. P. M. A. Crooijmans, M. A. M. Groenen, Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* **16**, 330 (2015).

7. T. Heinsohn, Animal translocation: long-term human influences on the vertebrate zoogeography of Australasia (natural dispersal versus ethnophoresy). *Aust. Zool.* **32**, 351–376 (2003).

8. A. A. Oktaviana, R. Joannes-Boyau, B. Hakim, B. Burhan, R. Sardi, S. Adhityatama, Hamrullah, I. Sumantri, M. Tang, R. Lebe, I. Ilyas, A. Abbas, A. Jusdi, D. E. Mahardian, S. Noerwidi, M. N. R. Ririmasse, I. Mahmud, A. Duli, L. M. Aksa, D. McGahan, P. Setiawan, A. Brumm, M. Aubert, Narrative cave art in Indonesia by 51,200 years ago. *Nature* **631**, 814–818 (2024).

9. M. Aubert, A. Brumm, M. Ramli, T. Sutikna, E. W. Saptomo, B. Hakim, M. J. Morwood, G. D. van den Bergh, L. Kinsley, A. Dosseto, Pleistocene cave art from Sulawesi, Indonesia. *Nature* **514**, 223–227 (2014).

10. G. Larson, T. Cucchi, M. Fujita, E. Matisoo-Smith, J. Robins, A. Anderson, B. Rolett, M. Spriggs, G. Dolman, T.-H. Kim, N. T. D. Thuy, E. Randi, M. Doherty, R. A. Due, R. Bollt, T. Djubiantono, B. Griffin, M. Intoh, E. Keane, P. Kirch, K.-T. Li, M. Morwood, L. M. Pedriña, P. J. Piper, R. J. Rabett, P. Shooter, G. Van den Bergh, E. West, S. Wickler, J. Yuan, A. Cooper, K. Dobney, Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4834–4839 (2007).

11. G. D. van den Bergh, H. J. M. Meijer, R. Due Awe, M. J. Morwood, K. Szabó, L. W. van den Hoek Ostende, T. Sutikna, E. W. Saptomo, P. J. Piper, K. M. Dobney, The Liang Bua faunal remains: a 95k.yr. sequence from Flores, East Indonesia. *J. Hum. Evol.* **57**, 527–537 (2009).

12. T. Sutikna, M. W. Tocheri, J. T. Faith, Jatmiko, R. Due Awe, H. J. M. Meijer, E. Wahyu Saptomo, R. G. Roberts, The spatio-temporal distribution of archaeological and faunal finds at Liang Bua (Flores, Indonesia) in light of the revised chronology for Homo floresiensis. *J. Hum. Evol.* **124**, 52–74 (2018).

13. P. Bellwood, *First Farmers* (Wiley-Blackwell, Hoboken, NJ, 2023).

14. F. Petchey, M. Spriggs, S. Bedford, F. Valentin, The chronology of occupation at Teouma, Vanuatu: Use of a modified chronometric hygiene protocol and Bayesian modeling to evaluate midden remains. *J. Archaeol. Sci. Rep.* **4**, 95–105 (2015).

15. *Lapita in the Southwest Pacific: Origins, Distribution, Chronology, Economy, and Transformation*.

16. F. Valentin, H. R. Buckley, E. Herrscher, R. Kinaston, S. Bedford, M. Spriggs, S. Hawkins, K. Neal, Lapita subsistence strategies and food consumption patterns in the community of Teouma (Efate, Vanuatu). *J. Archaeol. Sci.* **37**, 1820–1829 (2010).

17. G. Larson, R. Liu, X. Zhao, J. Yuan, D. Fuller, L. Barton, K. Dobney, Q. Fan, Z. Gu, X.-H. Liu, Y. Luo, P. Lv, L. Andersson, N. Li, Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7686–7691 (2010).

18. T. Denham, Early farming in Island Southeast Asia: an alternative hypothesis. *Antiquity* **87**, 250–257 (2013).

19. J. K. N. Layos, C. J. P. Godinez, L. M. Liao, Y. Yamamoto, J. S. Masangkay, H. Mannen, M. Nishibori, Origin and demographic history of Philippine pigs inferred from mitochondrial DNA. *Front. Genet.* **12** (2022).

20. J. B. Banayo, K. L. V. Manese, A. J. Salces, T. Yamagata, Phylogeny and genetic diversity of Philippine native pigs (Sus scrofa) as revealed by mitochondrial DNA analysis. *Biochem. Genet.* **61**, 1401–1417 (2023).

21. L. A. F. Frantz, J. G. Schraiber, O. Madsen, H.-J. Megens, A. Cagan, M. Bosse, Y. Paudel, R. P. M. A. Crooijmans, G. Larson, M. A. M. Groenen, Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148 (2015).

22. L. A. F. Frantz, J. Haile, A. T. Lin, A. Scheu, C. Geörg, N. Benecke, M. Alexander, A. Linderholm, V. E. Mullin, K. G. Daly, V. M. Battista, M. Price, K. J. Gron, P. Alexandri, R.-M. Arbogast, B. Arbuckle, A. Bălăşescu, R. Barnett, L. Bartosiewicz, G. Baryshnikov, C. Bonsall, D. Borić, A. Boroneanţ, J. Bulatović, C. Çakirlar, J.-M. Carretero, J. Chapman, M. Church, R. Crooijmans, B. D. Cupere, C. Detry, V. Dimitrijevic, V. Dumitraşcu, L. du Plessis, C. J. Edwards, C. M. Erek, A. Erim-Özdoğan, A. Ervynck, D. Fulgione, M. Gligor, A. Götherström, L. Gourichon, M. A. M. Groenen, D. Helmer, H. Hongo, L. K. Horwitz, E. K. Irving-Pease, O. Lebrasseur, J. Lesur, C. Malone, N. Manaseryan, A. Marciniak, H. Martlew, M. Mashkour, R. Matthews, G. M. Matuzeviciute, S. Maziar, E. Meijaard, T. McGovern, H.-J. Megens, R. Miller, A. F. Mohaseb, J. Orschiedt, D. Orton, A. Papathanasiou, M. P. Pearson, R. Pinhasi, D. Radmanović, F.-X. Ricaut, M. Richards, R. Sabin, L. Sarti, W. Schier, S. Sheikhi, E. Stephan, J. R. Stewart, S. Stoddart, A. Tagliacozzo, N. Tasić, K. Trantalidou, A. Tresset, C. Valdiosera, Y. van den Hurk, S. Van Poucke, J.-D. Vigne, A. Yanevich, A. Zeeb-Lanz, A. Triantafyllidis, M. T. P. Gilbert, J. Schibler, P. Rowley-Conwy, M. Zeder, J. Peters, T. Cucchi, D. G. Bradley, K. Dobney, J. Burger, A. Evin, L. Girdland-Flink, G. Larson, Ancient pigs reveal a near-complete genomic turnover following their introduction to Europe. *Proceedings of the National Academy of Sciences* **116**, 17231–17238 (2019).

23. C. P. Groves, *Ancestors for the Pigs : Taxonomy and Phylogeny of the Genus Sus* (Australian National University Press, Canberra, 1981).

24. Mammals Collections Search. https://collections.nmnh.si.edu/search/mammals/?ark=ark:/65665/36beb92ac78e0438991aff7af501110a7.

25. IUCN, Sus celebensis: Burton, J., Mustari, A. & Rejeki, I, IUCN (2016); https://doi.org/10.2305/iucn.uk.2020-2.rlts.t41773a44141588.en.

26. M. Chintalapati, N. Patterson, P. Moorjani, The spatiotemporal patterns of major human admixture events during the European Holocene. *Elife* **11**, e77625 (2022).

27. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).

28. S. Rubinacci, D. M. Ribeiro, R. J. Hofmeister, O. Delaneau, Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* **53**, 120–126 (2021).

29. H. Hilmarsson, A. S. Kumar, R. Rastogi, C. D. Bustamante, D. M. Montserrat, A. G. Ioannidis, High resolution ancestry deconvolution for next generation genomic data, *bioRxiv* (2021). https://doi.org/10.1101/2021.09.19.460980.

30. P. Skoglund, C. Posth, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G. R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M. P. Cox, F. R. Friedlaender, J. S. Friedlaender, T. Kivisild, G. Koki, P. Kusuma, D. A. Merriwether, F.-X. Ricaut, J. T. S. Wee, N. Patterson, J. Krause, R. Pinhasi, D. Reich, Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).

31. B. Egloff, Recent prehistory in southeast Papua. (1979).

32. W. B. Masse, "The archaeology and ecology of fishing in the Belau Islands, Micronesia," thesis, Southern Illinois University (1989).

33. G. Clark, C. Reepmeyer, Last millennium climate change in the occupation and abandonment of Palau's Rock Islands. *Archaeol. Ocean.* **47**, 29–38 (2012).

34. P. Bellwood, E. Dizon, "The batanes islands, their first observers, and previous archaeology" in *4000 Years of Migration and Cultural Exchange (Terra Australis 40): The Archaeology of the Batanes Islands, Northern Philippines* (ANU Press, 2013), pp. 1–8.

35. P. J. Reimer, W. E. N. Austin, E. Bard, A. Bayliss, P. G. Blackwell, C. Bronk Ramsey, M. Butzin, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, I. Hajdas, T. J. Heaton, A. G. Hogg, K. A. Hughen, B. Kromer, S. W. Manning, R. Muscheler, J. G. Palmer, C. Pearson, J. van der Plicht, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, C. S. M. Turney, L. Wacker, F. Adolphi, U. Büntgen, M. Capano, S. M. Fahrni, A. Fogtmann-Schulz, R. Friedrich, P. Köhler, S. Kudsk, F. Miyake, J. Olsen, F. Reinig, M. Sakamoto, A. Sookdeo, S. Talamo, The IntCal20 Northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).

36. C. Ottoni, L. G. Flink, A. Evin, C. Geörg, B. De Cupere, W. Van Neer, L. Bartosiewicz, A. Linderholm, R. Barnett, J. Peters, R. Decorte, M. Waelkens, N. Vanderheyden, F.-X. Ricaut, C. Cakirlar, O. Cevik, A. R. Hoelzel, M. Mashkour, A. F. M. Karimlu, S. S. Seno, J. Daujat, F. Brock, R. Pinhasi, H. Hongo, M. Perez-Enciso, M. Rasmussen, L. Frantz, H.-J. Megens, R. Crooijmans, M. Groenen, B. Arbuckle, N. Benecke, U. S. Vidarsdottir, J. Burger, T. Cucchi, K. Dobney, G. Larson, Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol. Biol. Evol.* **30**, 824–832 (2013).

37. P. B. Damgaard, A. Margaryan, H. Schroeder, L. Orlando, E. Willerslev, M. E. Allentoft, Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015).

38. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).

39. C. Carøe, S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, M. T. P. Gilbert, Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).

40. S. A. Miller, D. D. Dykes, H. F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).

41. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

42. M.-T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).

43. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

44. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

45. J. A. Fellows Yates, T. C. Lamnidis, M. Borry, A. Andrades Valtueña, Z. Fagernäs, S. Clayton, M. U. Garcia, J. Neukamm, A. Peltzer, Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* **9**, e10947 (2021).

46. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

47. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

48. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

49. K. Katoh, J. Rozewicki, K. D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).

50. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

51. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

52. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

53. H. P. Eggertsson, S. Kristmundsdottir, D. Beyter, H. Jonsson, A. Skuladottir, M. T. Hardarson, D. F. Gudbjartsson, K. Stefansson, B. V. Halldorsson, P. Melsted, GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nat. Commun.* **10**, 5402 (2019).

54. M. Johnsson, A. Whalen, R. Ros-Freixedes, G. Gorjanc, C.-Y. Chen, W. O. Herring, D.-J. de Koning,

J. M. Hickey, Genetic variation in recombination rate in the pig. *Genet. Sel. Evol.* **53**, 54 (2021).

55. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

56. J. Meisner, S. Liu, M. Huang, A. Albrechtsen, Large-scale inference of population structure in presence of missingness using PCA. *Bioinformatics* **37**, 1868–1875 (2021).

57. M. Bosse, H.-J. Megens, L. A. F. Frantz, O. Madsen, G. Larson, Y. Paudel, N. Duijvesteijn, B. Harlizius, Y. Hagemeijer, R. P. M. A. Crooijmans, M. A. M. Groenen, Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392 (2014).

58. P. Librado, L. Orlando, Struct-f4: a Rcpp package for ancestry profile and population structure inference from f4-statistics. *Bioinformatics* **38**, 2070–2071 (2022).

59. S. White, From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History. *Environ. Hist. Durh. N. C.* **16**, 94–120 (2011).

60. A. Bergström, D. W. G. Stanton, U. H. Taron, L. Frantz, M.-H. S. Sinding, E. Ersmark, S. Pfrengle, M. Cassatt-Johnstone, O. Lebrasseur, L. Girdland-Flink, D. M. Fernandes, M. Ollivier, L. Speidel, S. Gopalakrishnan, M. V. Westbury, J. Ramos-Madrigal, T. R. Feuerborn, E. Reiter, J. Gretzinger, S. C. Münzel, P. Swali, N. J. Conard, C. Carøe, J. Haile, A. Linderholm, S. Androsov, I. Barnes, C. Baumann, N. Benecke, H. Bocherens, S. Brace, R. F. Carden, D. G. Drucker, S. Fedorov, M. Gasparik, M. Germonpré, S. Grigoriev, P. Groves, S. T. Hertwig, V. V. Ivanova, L. Janssens, R. P. Jennings, A. K. Kasparov, I. V. Kirillova, I. Kurmaniyazov, Y. V. Kuzmin, P. A. Kosintsev, M. Lázničková-Galetová, C. Leduc, P. Nikolskiy, M. Nussbaumer, C. O'Drisceoil, L. Orlando, A. Outram, E. Y. Pavlova, A. R. Perri, M. Pilot, V. V. Pitulko, V. V. Plotnikov, A. V. Protopopov, A. Rehazek, M. Sablin, A. Seguin-Orlando, J. Storå, C. Verjux, V. F. Zaibert, G. Zazula, P. Crombé, A. J. Hansen, E. Willerslev, J. A. Leonard, A. Götherström, R. Pinhasi, V. J. Schuenemann, M. Hofreiter, M. T. P. Gilbert, B. Shapiro, G. Larson, J. Krause, L. Dalén, P. Skoglund, Grey wolf genomic history reveals a dual ancestry of dogs. *Nature* **607**, 313–320 (2022).

61. A. Evin, K. Dobney, R. Schafberg, J. Owen, U. S. Vidarsdottir, G. Larson, T. Cucchi, Phenotype and animal domestication: A study of dental variation between domestic, wild, captive, hybrid and insular Sus scrofa. *BMC Evol. Biol.* **15**, 6 (2015).

62. F. J. Rohlf, tpsDig, version 2.10. (2006).

63. S. Schlager, "Chapter 9 - Morpho and Rvcg – Shape Analysis in R: R-Packages for Geometric Morphometrics, Shape Analysis and Surface Manipulations" in *Statistical Shape and Deformation Analysis*, G. Zheng, S. Li, G. Székely, Eds. (Academic Press, 2017), pp. 217–256.

64. A. Evin, L. G. Flink, A. Bălăşescu, D. Popovici, R. Andreescu, D. Bailey, P. Mirea, C. Lazăr, A. Boroneanţ, C. Bonsall, U. S. Vidarsdottir, S. Brehard, A. Tresset, T. Cucchi, G. Larson, K. Dobney, Unravelling the complexity of domestication: a case study using morphometrics and ancient DNA analyses of archaeological pigs from Romania. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130616 (2015).

Acknowledgements: We are grateful to Mark Golitko and John Edward Terrell for support acquiring material used in this study.

Funding: L.A.F.F., G.L., D.W.G.S., A.M., and K.D.were supported by either European Research Council grants (ERC-2013-StG-337574-UNDEAD and ERC-2019-StG-853272-PALAEOFARM) or Natural

Environmental Research Council grants (NE/K005243/1 and NE/K003259/1) or both. HPC computing was performed on the BioHPC (DFG INST 86/2050-1 FUGG) and Linux-Cluster of the Leibniz Supercomputing Centre (LRZ Munich). L.A.F.F. and R.D. were funded by the NERC-Ristekdikti Newton Wallacea joint research programme (grant number NE/S007067/1; Ristekdikti Grant No: NKB-1799/UN2.R3.1/HKP.05.00/2019 "Biodiversity, Environmental Change, and Land-use Policy in Sulawesi and Maluku").

Author contributions:

Conceptualization: D.W.G.S., A.M., A.E., K.T., A.L., R.D., E.M., P.Be., T.H., P.J.P., L.Sca., M.P.C., T.D., K.D., G.L., L.F.; Formal Analysis: D.W.G.S., A.M., A.E., K.T., S.I.N., T.C., U.S.V., L.F.; Investigation: A.M., K.T., A.L., R.D., O.T., L.G-F., H.H., C.L., K.J.M., C.M., K.N., R.P., P.S., J.G.; Resources: N.A., A.A., P.Ba., S.B., P.Be., A.B., T.C.T., G.C., M.E., B.H., S.H., H.H., K.M.H., M.J.H., M.J.L.A.B., E.Z.D., H-C.H., K.G., A.C.K., C.L., A.A.M., E.M., K.J.M., C.M., K.M., K.N., M.O., R.P., K.S., L.Sch., P.S., M.S., S.T.T., J.G.; Data Curation: L.F.; Writing – Original Draft: D.W.G.S., A.M., A.E., K.T., A.L., K.D., G.L., L.F.; Writing – Review & Editing: D.W.G.S., A.M., A.E., K.T., A.L., R.D., O.T., S.I.N., N.A., A.A., R.B., P.Ba., S.B., P.Be., A.B., T.C.T., G.C., R.C., T.C., M.E., L.G-F., P.G., M.G., B.H., S.H., H.H., K.M.H., M.J.H., M.J.L.A.B., E.Z.D., H-C.H., K.G., T.H., A.C.K., C.L., A.A.M., H-J.M., E.M., K.J.M., C.M., K.M., K.N., M.O., R.P., P.J.P., L.Sca., K.S., L.Sch., P.S., M.S., S.T.T., U.S.V., M.P.C., T.D., J.G., K.D., G.L., L.F.; Visualization: D.W.G.S., A.M., A.E., K.T., S.I.N., L.F.; Supervision: T.C., H-C.H., J.G., K.D., G.L., L.F.; Funding acquisition: E.M., K.D., G.L., L.F.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability:

All genomics data have been deposited at the European Nucleotide Archive (ENA) with project number PRJEB83975. All morphometric data is available as in Data S3.

Supplementary Materials

Materials and Methods
Supplementary Text
Figs. S1 to S15
Data S1 and S3
References (*31-64*)

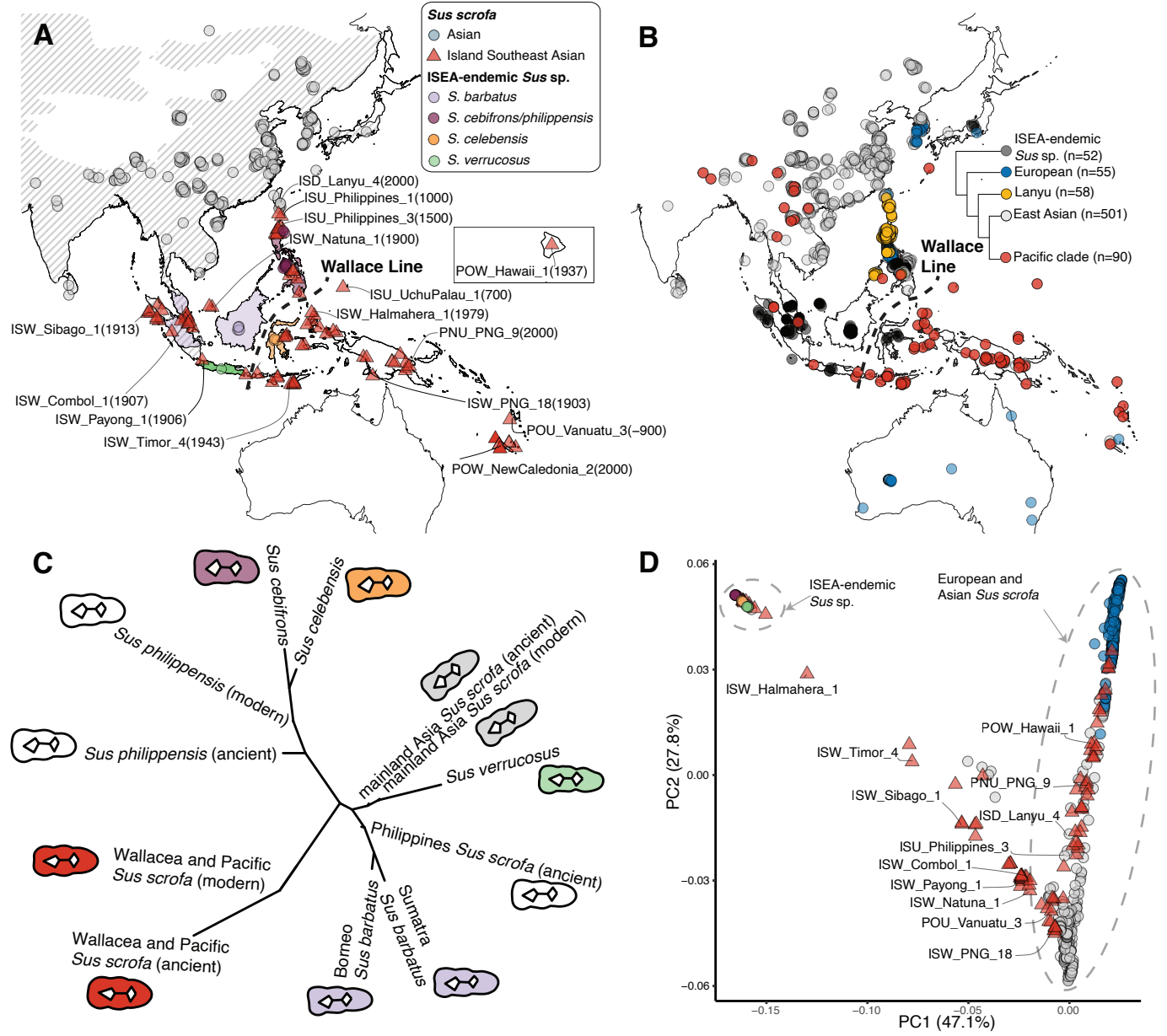
Figures:

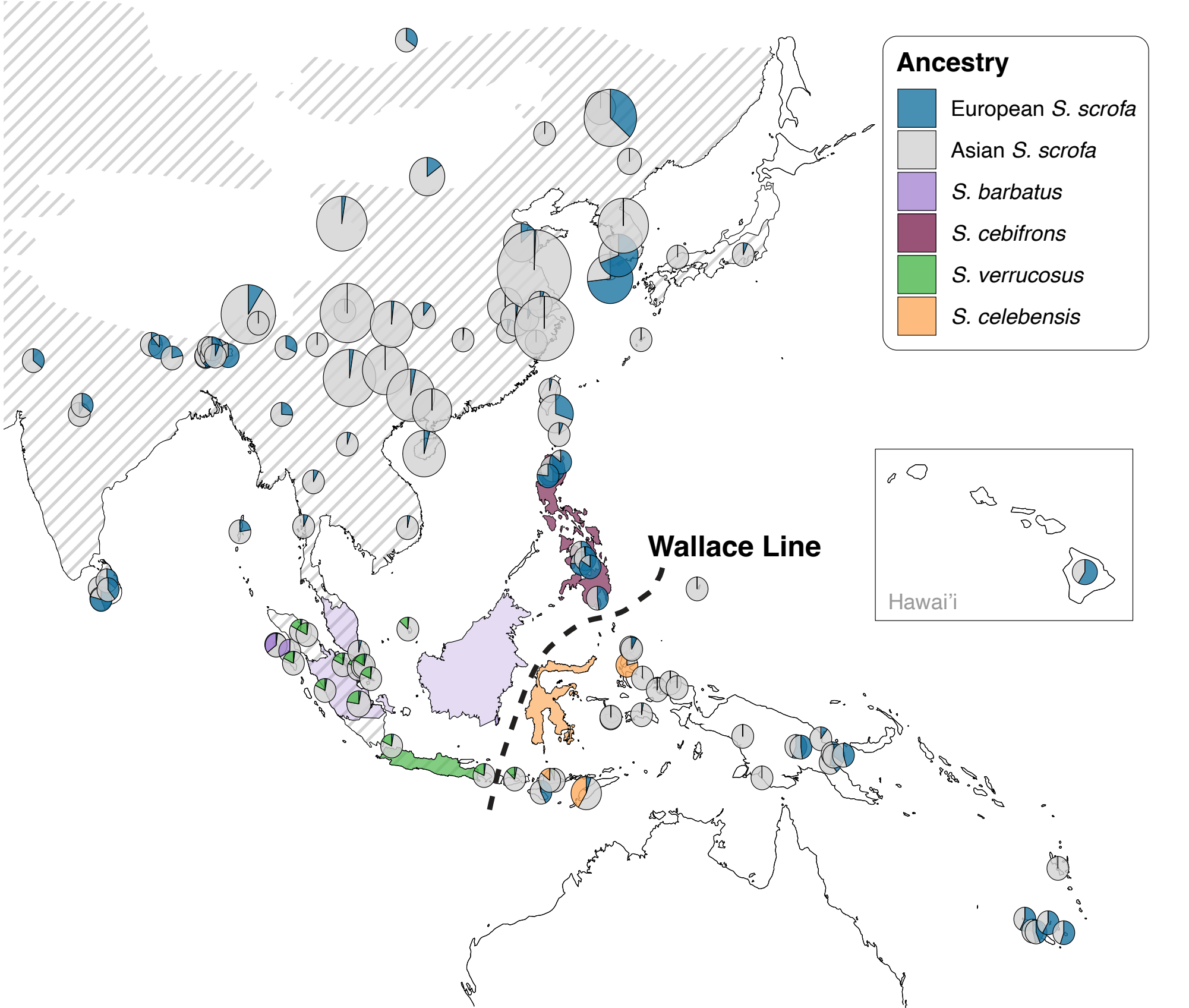
Fig. 1 A. Map of sampling locations. Red triangles indicate 49 newly sequenced historical *S. scrofa* specimens (dating from 1871-1994) from ISEA (cataloged as *S. scrofa* in museum collections) and three archeological specimens from the Philippines (~1500 AD), Palau (Micronesia; ~672 AD), and the Lapita culture site of Teouma on Efate Island, Vanuatu. Gray dots represent individual *Sus scrofa* genomes (some of which are newly sequenced; see Data S1) from mainland Asia (both South Asia and East Asia). The locations of the ISEA endemic *Sus* species individual genome analyzed in this study are noted in purple (*S. barbatus*; Borneo), cream (*S. cebifron*s), orange (*S. celebensis*), and green (*S. verrucosus*). **B. Mitochondrial analysis.** Geographic distribution of major mtDNA haplogroups based on data from ~705 control regions and ~323 whole mitochondrial genome data. The tree on the left is a schematic representation illustrating the phylogenetic relationships between these major haplogroups. The complete phylogenetic tree, based on whole mitochondrial genome data, is available in Fig. S3. This figure illustrates the distribution of the red dots, representing the "Pacific Clade," extending west of the Wallace Line into Western Indonesia, the Philippines, and mainland Asia (as shown in the full mitogenome tree). **C. Morphometric analysis.** This neighbour-joining network illustrates the morphometric similarities in lower third molar (m3) shape among *Sus scrofa* and ISEA endemic *Sus* species, based on Mahalanobis distances. The network was generated following a discriminant analysis that utilized the first 15 principal components from a preceding Principal Component Analysis (PCA). These selected components explain 88.7% of the total shape variance and were chosen to maximize the morphological distinctions between the groups. The tooth outlines shown represent the mean shape for each population, derived from separate superimposition analyses. **D. Principal component analysis (PCA).** This PCA (see panel A for label), based on pseudohaploid genotype from >500 individuals, illustrates the genetic variation among Eurasian *S. scrofa* and ISEA-endemic *Sus* species. The first principal component (PC1) clearly separates the ISEA-endemic *Sus* species from *S. scrofa* individuals, while the second principal component (PC2) distinguishes between European and Asian *S. scrofa*. The red triangles (*S. scrofa* genomes from ISEA) are scattered across the PC1-2 space, reflecting their complex ancestry, which includes contributions from Asian and European *S. scrofa* as well as ISEA-endemic *Sus* ancestry.

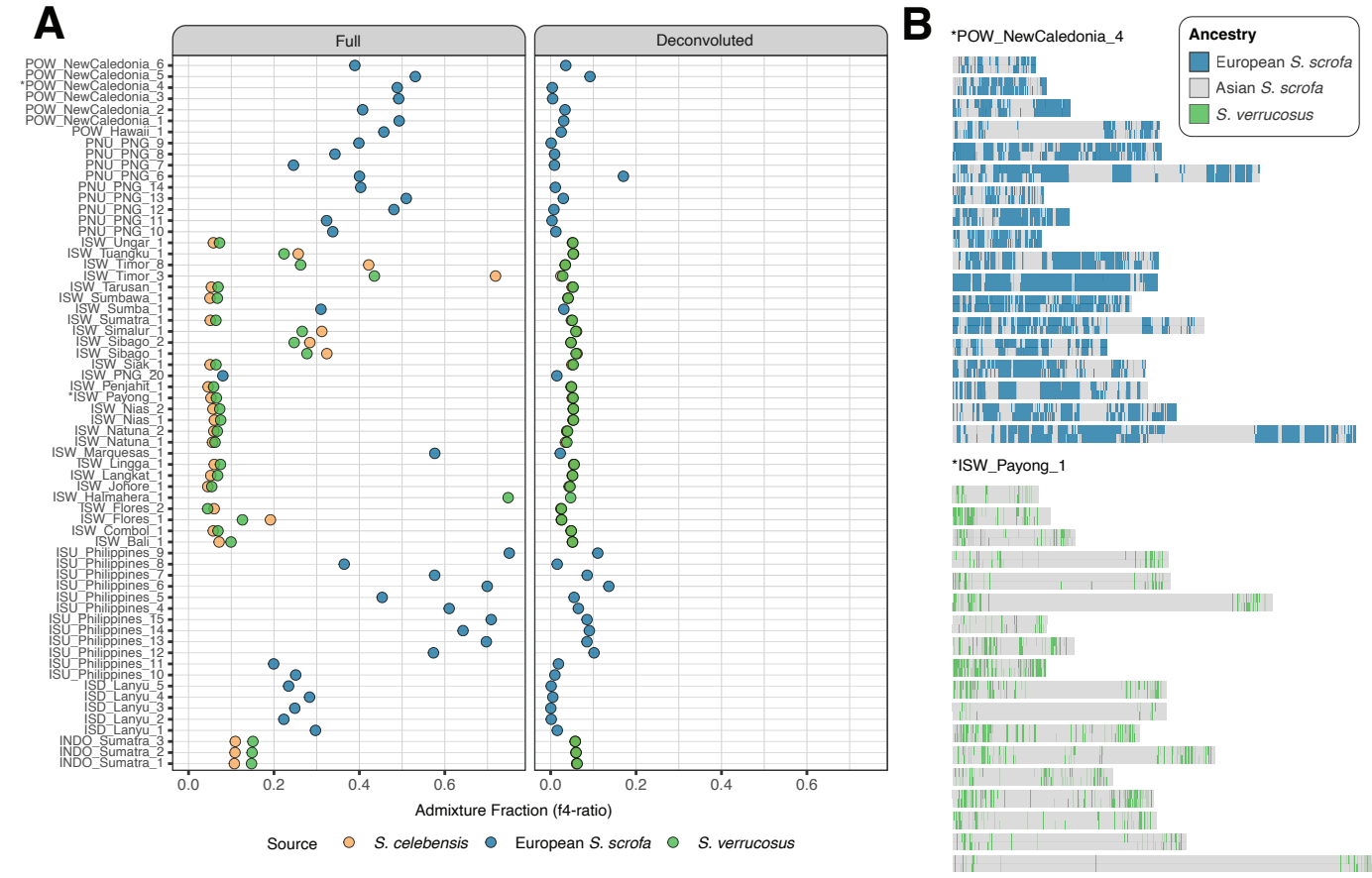
Fig. 2. Ancestry of pigs inferred using ADMIXTURE and D-statistics. The ancestry of each individual Asian pig in our dataset was decomposed into five components using a combination of supervised ADMIXTURE analysis and D-statistics. Initially, the ancestry was decomposed with K=3 in ADMIXTURE (see Fig. S6), representing European *S. scrofa*, Asian *S. scrofa*, and ISEA-endemic *Sus* ancestry. The ISEA-endemic *Sus* ancestry component was further resolved using D-statistics (see Supplementary Material). Each pie chart on the figure represents a sampling location, and its size is proportional to the number of samples from that location.

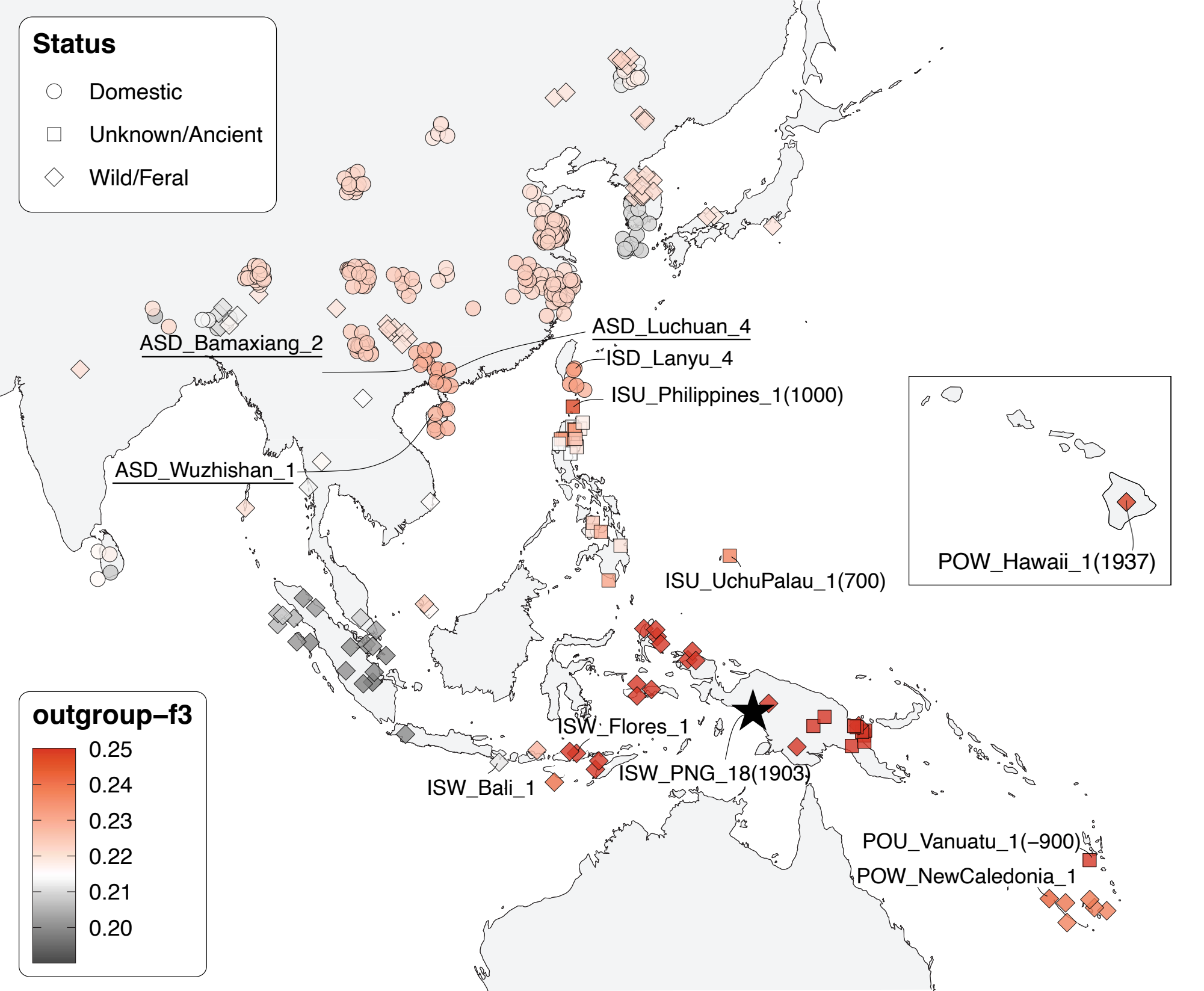
Fig. 3A. **Accuracy of ancestry deconvolution.** European *S. scrofa* / ISEA-endemic *Sus* ancestry proportion in genomes of mixed ISEA individuals before (full) and after (deconvoluted) ancestry, with deconvolution calculated using the f4 ratio (see Supplementary Material). Values close to 0 in the right panel (deconvoluted) indicate successful deconvolution, meaning the European *S. scrofa* / ISEA-endemic *Sus* ancestry was effectively removed from the genome. **B.** Blocks of European *S. scrofa* and *S. verrucosus* ancestry across chromosomes from a New Caledonian wild-caught individual and a wild pig from Payong, a small island a few km north of Java (Fig. 1A). This figure illustrates that the European *S. scrofa* ancestry blocks in the New Caledonian pig genome are considerably larger than the *S. verrucosus* ancestry blocks in the Payong pig genome. This difference indicates a more recent admixture event in the New Caledonian pig (dated to ~1922 using DATES (*26*); see Supplementary Material).

Fig. 4. Outgroup-f3 after deconvolution. Outgroup-f3 (shared drift) of the form (Pygmy hog [outgroup], X, ISW_PNG_18) plotted on a map, where X represents any Asian individual in our dataset. Prior to this analysis, European *S. scrofa* and ISEA endemic *Sus* ancestry was removed from the genomes of admixed individuals using ancestry deconvolution. ISW_PNG_18 was selected as a proxy for the approximately 2,900-year-old Vanuatu pig specimen excavated from a Lapita culture archaeological site (see Fig. S11). Higher outgroup-f 3 values indicate a greater degree of genetic affinity with the ISW_PNG_18 individual, which serves as a representative of pig ancestry associated with the dispersal of Austronesian language speakers. This figure demonstrates a strong genetic affinity between pigs east of the Wallace Line and those in mainland Asia (i.e. those underlined) to those introduced by Austronesian language speakers, rather than with pigs from west of the Wallace Line in ISEA.


A
Sus scrofa
Asian
Island Southeast Asian
ISEA-endemic Sus sp.
S. barbatus
S. cebifrons/philippensis
S. celebensis
S. verrucosus
ISD_Lanyu_4(2000)
ISU_Philippines_1(1000)
ISU_Philippines_3(1500)
ISW_Natuna_1(1900)
Wallace Line
POW_Hawaii_1(1937)
ISU_UchuPalau_1(700)
ISW_Halmahera_1(1979)
PNU_PNG_9(2000)
ISW_Sibago_1(1913)
ISW_Combol_1(1907)
ISW_Payong_1(1906)
ISW_Timor_4(1943)
ISW_PNG_18(1903)
POU_Vanuatu_3(~900)
POW_NewCaledonia_2(2000)
B
ISEA-endemic Sus sp. (n=52)
European (n=55)
Lanyu (n=58)
East Asian (n=501)
Pacific clade (n=90)
Wallace Line
C
Sus cebifrons
Sus celebensis
Sus philippensis (modern)
Sus philippensis (ancient)
mainland Asia Sus scrofa (ancient)
mainland Asia Sus scrofa (modern)
Sus verrucosus
Philippines Sus scrofa (ancient)
Wallacea and Pacific Sus scrofa (modern)
Wallacea and Pacific Sus scrofa (ancient)
Borneo Sus barbatus
Sumatra Sus barbatus
D
ISEA-endemic Sus sp.
European and Asian Sus scrofa
ISW_Halmahera_1
POW_Hawaii_1
ISW_Timor_4
PNU_PNG_9
ISW_Sibago_1
ISD_Lanyu_4
ISU_Philippines_3
ISW_Combol_1
ISW_Payong_1
ISW_Natuna_1
POU_Vanuatu_3
ISW_PNG_18
PC2 (27.8%)
PC1 (47.1%)
0.06
0.03
0.00
−0.03
−0.06
−0.15
−0.10
−0.05
0.00


Ancestry
European S. scrofa
Asian S. scrofa
S. barbatus
S. cebifrons
S. verrucosus
S. celebensis
Wallace Line
Hawai'i


A
Full
Deconvoluted
POW_NewCaledonia_6
POW_NewCaledonia_5
*POW_NewCaledonia_4
POW_NewCaledonia_3
POW_NewCaledonia_2
POW_NewCaledonia_1
POW_Hawaii_1
PNU_PNG_9
PNU_PNG_8
PNU_PNG_7
PNU_PNG_6
PNU_PNG_14
PNU_PNG_13
PNU_PNG_12
PNU_PNG_11
PNU_PNG_10
ISW_Ungar_1
ISW_Tuangku_1
ISW_Timor_8
ISW_Timor_3
ISW_Tarusan_1
ISW_Sumbawa_1
ISW_Sumba_1
ISW_Sumatra_1
ISW_Simalur_1
ISW_Sibago_2
ISW_Sibago_1
ISW_Siak_1
ISW_PNG_20
ISW_Penjahit_1
*ISW_Payong_1
ISW_Nias_2
ISW_Nias_1
ISW_Natuna_2
ISW_Natuna_1
ISW_Marquesas_1
ISW_Lingga_1
ISW_Langkat_1
ISW_Johore_1
ISW_Halmahera_1
ISW_Flores_2
ISW_Flores_1
ISW_Combol_1
ISW_Bali_1
ISU_Philippines_9
ISU_Philippines_8
ISU_Philippines_7
ISU_Philippines_6
ISU_Philippines_5
ISU_Philippines_4
ISU_Philippines_15
ISU_Philippines_14
ISU_Philippines_13
ISU_Philippines_12
ISU_Philippines_11
ISU_Philippines_10
ISD_Lanyu_5
ISD_Lanyu_4
ISD_Lanyu_3
ISD_Lanyu_2
ISD_Lanyu_1
INDO_Sumatra_3
INDO_Sumatra_2
INDO_Sumatra_1
0.0
0.2
0.4
0.6
Admixture Fraction (f4-ratio)
Source
S. celebensis
European S. scrofa
S. verrucosus
B
*POW_NewCaledonia_4
*ISW_Payong_1
Ancestry
European S. scrofa
Asian S. scrofa
S. verrucosus


Status
Domestic
Unknown/Ancient
Wild/Feral
ASD_Bamaxiang_2
ASD_Luchuan_4
ISD_Lanyu_4
ISU_Philippines_1(1000)
ASD_Wuzhishan_1
ISU_UchuPalau_1(700)
POW_Hawaii_1(1937)
ISW_Flores_1
ISW_PNG_18(1903)
ISW_Bali_1
POU_Vanuatu_1(−900)
POW_NewCaledonia_1
outgroup−f3
0.25
0.24
0.23
0.22
0.21
0.20

Supplementary Materials for

Genomic and morphometric evidence for Austronesian-mediated pig translocation in the Pacific

David W. G. Stanton[1,2] †, Aurelie Manin[3,4] †, Allowen Evin[5] †, Kristina Tabbada[3] †, Anna Linderholm[6,7,3] †, Rosie Drinkwater[1,8], Olaf Thalmann[8], Said I. Ng'ang'a[1,8], Noel Amano[9], Atholl Anderson[10], Ross Barnett[11], Patrick Barrière[12], Stuart Bedford[13,14], Peter Bellwood[13], Adam Brumm[15,16], Trung Cao Tien[17,18], Geoffrey Clark[10], Richard Crooijmans[19], Thomas Cucchi[20], Michelle Eusebio[21], Linus Girdland-Flink[22], Peter Galbusera[23], Martien Groenen[19], Budianto Hakim[24], Stuart Hawkins[13], Holly Heiniger[25,26], Kristofer M. Helgen[27], Michael J. Herrera[21], Mary Jane Louise A. Bolunia[28], Eusebio Z. Dizon[28] ‡, Hsiao-chun Hung[10], Karen Greig[29], Terry Hunt[30], Andrew C. Kitchener[31,32], Carol Lee[33,34], Alastair A. Macdonald[35], Hendrik-Jan Megens[19], Erik Meijaard[36,37], Kieren J. Mitchell[25,26], Christopher Moran[33], Karen Mudar[38], Karma Nidup[33], Marc Oxenham[13,39], Rinzin Pem[40], Philip J. Piper[13], Lachie Scarsbrook[3,8], Kyle Schachtschneider[41], Lawrence Schook[42], Pradeepa Silva[43], Matthew Spriggs[13,44], Samuel T. Turvey[45], Una S. Viðarsdóttir[46], Murray P. Cox[47,48], Tim Denham[13], Jaime Gongora[33], Keith Dobney[22,49], Greger Larson[3] *, Laurent Frantz[1,8] *

Corresponding author: laurent.frantz@lmu.de; greger.larson@arch.ox.ac.uk

The PDF file includes:
Materials and Methods
Supplementary Text
Figs. S1 to S16

Other Supplementary Materials for this manuscript include the following:
Data S1 to S3

Materials and Methods

Archeological specimens and site description of DNA samples

Uripiv, Vanuatu (2900-2000 BP; POU_Vanuatu_3; contact: Stuart Bedford, School of Archaeology and Anthropology, College of Arts and Social Sciences, Australian National University, Canberra, ACT, Australia)

The small islands of Vao and Uripiv are located off the northeast coast of Malekula Island in northern Vanuatu. Excavations on uplifted beach terraces on the sheltered coast of the islands revealed dense concentrations of midden deposits associated with the colonising Lapita phase (c. 2900 BP) through to c. 2000 BP. The sites are rich in faunal remains, both terrestrial and marine, shellfish, pottery and other artefacts. The sites on both islands were deeply stratified, up to 1.8 m deep to the sterile beach sand in some places, and generally very well preserved showing limited disturbance.

Teouma, Vanuatu (3000-2600 BP; POU_Vanuatu_4, POU_Vanuatu_5; contact: Stuart Bedford, School of Archaeology and Anthropology, College of Arts and Social Sciences, Australian National University, Canberra, ACT, Australia)

The Teouma Lapita site is located on the south coast of Efate Island, Vanuatu, and is currently 800 metres from Teouma Bay. Tectonic uplift, volcanic ashfall prior to and during the period of human utilisation of the site, and alluvial deposition from an adjacent stream have altered its immediate coastal location during the Lapita period at c. 3000 BP. Excavations by a joint ANU-Vanuatu National Museum team took place between 2004 and 2010, during which 68 burial features, including remains of just over 100 individuals were found. Teouma has numerous indicators of being an initial colonisation site for Efate, including extinct faunal remains, early ceramic forms and decoration, and exotic New Britain obsidian. This combined with the extensive and previously very rare Lapita skeletal remains underline its significance to those investigating colonisation in this region. There is evidence of continuing occupation at the site to c. 2600 BP. During the Lapita phase the site was composed of a cemetery and a contemporary settlement with a midden deposit. During later phases settlement expanded across the cemetery area.

Wanigela, Papua New Guinea (600-700 BP; PNU_PNG_2, PNU_PNG_3; contact: Geoffrey Clark, Archaeology and Natural History, School of Culture, History and Language, Australian National University, Canberra, Australia)

The 14C sample was an adult pig tooth and fragment of mandible from Oreresan Village, Mound B (Spit 8 mixed) dated by Wk-30430, CRA 675 +/- 30, 13C -20.1, C:N 3.3. All the aDNA samples are from the Mound B excavation which is located in Oreresan Village located in coastal Eastern New Guinea in Collingwood Bay. The mound was excavated by Brian Egloff in the 1960s and site details are in his thesis publication (*31*).

Mound B is an oval-shaped mound near the Sasap River that is around 40 m in length and 17 m in width with a maximum height of 1.4 m. Excavations at the site totalled 22 m2 with sediments divided into eight zones and subzones – Zone I (top), Zone IIA, IIB, IIC, Zone III, Zone IVA, IVB (base). Pig bones were common in all excavation levels and comprised 80% of the identifiable vertebrate remains (*31*). Many pigs had been killed before their third molars had erupted when the body weight was estimated to be less than 80-120 kg. Four radiocarbon results obtained by Egloff and suggested that the Mound B materials were deposited between AD 100 and AD 1450. The large amount of pig remains in Mound B might represent debris from ceremonial feasting indicating that the importance of pigs in traditional societies in New Guinea dates to AD 1300-1400 (Calib 7.10 result for Wk-30430 at 95.4% is AD 1270-1390).

Uchul A Rois, Palau (900-1450 AD; ISU_UchuPalau_1, ISU_UchuPalau_2; contact: Geoffrey Clark, Archaeology and Natural History, School of Culture, History and Language, Australian National University, Canberra, Australia)

Excavations of ~4 m2 in Uchularois Cave by Masse (1989) (*32*)recovered an assemblage of pig bones totalling 110 identified specimens. Radiocarbon dates from Uchularois Cave have been recently revised and suggest that midden deposition began at AD 900, with site abandonment by AD 1450 (*33*). ISU_UchuPalau_1 was radiocarbon dated to ~1,300 calAD (OxA-44064).

Pamayan, Philippines (1420–1630 AD; ISU_Philippines_3; contact: Peter Bellwood, School of Archaeology and Anthropology, College of Arts and Social Sciences, Australian National University, Canberra, ACT, Australia)

Pamayan, Savidug village, Savidug Island, Batanes Islands, northern Philippines (Late prehistoric, overlapping with early Spanish period). The Pamayan shell midden lies at the back of Savidug village, exposed by a road cutting along the base of a limestone hill. A surface presence of red-slipped pottery indicated the existence of a potentially interesting site, so two small excavation sections of 60 by 80 cm were excavated 16 m apart back into the section exposed in the road cutting, exposing shell midden to a maximum depth of 105 cm beneath the topsoil (*34*). However, the recency of the Pamayan midden was rendered obvious by the occurrence of a few imported Chinese ceramic sherds down to 95 cm depth. A C14 date of 418±41 uncal. BP (WK 13091) from the base of the cultural layer indicates that the midden overlaps in its age of formation with the arrival of Spanish missionaries on the island, although there is a strong possibility that the archaeological materials have migrated downslope from a former site on top of the hill, a circumstance which makes the precise date of the pig bone sample uncertain. The pig bone was therefore dated using radiocarbon dating to 1475 calAD (OxA-43990).

Savidug Dune Site, Philippines (600-1000 BP; ISU_Philippines_1; contact: Peter Bellwood, School of Archaeology and Anthropology, College of Arts and Social Sciences, Australian National University, Canberra, ACT, Australia)

Savidug Dune Site, Savidug village, Savidug Island, Batanes Islands, northern Philippines. The Savidug Dune Site has two stratigraphic and cultural phases of usage (*34*). An upper cultural layer dated c.1000 to 600 BP, with imported ceramics and iron, hence overlapping in date with the formation of the Pamayan shell midden. A lower cultural layer dated c.3200 to 2000/1500 BP, which can be divided into an earlier and a later phase. The earlier phase is associated with rare examples of circle-stamped pottery, dated 3200 to 2700 BP in this site and from parallels at Sunget on Batan Island and Anaro on Itbayat Island. The later phase, dated c.2700 to 2000/1500 BP, is associated with jar burial and Taiwan nephrite.

Radiocarbon dating

Two samples were radiocarbon dated at the Oxford Radiocarbon Accelerator Unit (ORAU) The dates were calibrated on OxCal using the IntCal20 curve (*35*):

Sample ID	**Site, country**	**Uncalibrated radiocarbon age BP**	**Error (+/-)**	**Calibrated age BP**	**Calibration curve**	**Lab code**
ISU_UchuPalau_1	Uchul a Rois, Palau	1370	18	1308-1279	IntCal20	OxA-44064
ISU_Philippines_3	Pamayan, Philippines	424	16	511-475	IntCal20	OxA-43990

Museum specimens

Museum samples for DNA were collected from the AMNH (American Museum of Natural History, New York, USA), the Australian Museum (Sydney, Australia), FMNH (Field Museum of Natural History, Chicago, USA), the Long An Museum (Vietnam), the NHM (Natural History Museum, London, UK), the NMS (National Museum of Scotland, Edinburgh, Scotland), the Naturalis (Naturalis Leiden, The Netherlands), the OUM (Oxford University Museum, Oxford, UK), and the USNM (Smithsonian National Museum of Natural History, Washington DC, USA). More information, including their museum identification for these samples can be obtained from Data S1.

Archeological and Museum specimen extraction and library

DNA extraction

Samples were extracted in two different ancient DNA facilities - Durham University and the University of Oxford - using conventional set-up for ancient and degraded genomic material (strictly isolated pre- and post-PCR rooms, use of full PPE with overshoes and double set of gloves [nitrile and latex], systematic extraction and library blanks was added to each batch of 12 samples. All equipment and work surfaces were cleaned before and after each use with a dilute solution of bleach (5-10%) followed by water and ethanol (99%). Pipettes and plastic racks were UV-irradiated in a crosslinker (254 nm wavelength) prior to and after use and/or decontaminated using "DNA AWAY" (Fisher Scientific SAS) or bleach followed by purified water. In both laboratories the ancient pig remains were prepared for DNA extraction by removing approximately two a millimetres layer of the outer bone surface by abrasion using a Dremel drill with clean cut-off wheels (Dremel no 409), targeting compact cortical bone or dentine. The bone was then pulverised in a Micro-dismembrator (Sartorious-Stedim Biotech), followed by collection in 15 mL Grainer tubes (Durham) or in 2 mL ependorf tubes (Oxford). Milling containers and grinding balls were subsequently suspended and cleaned in 1% virkon (Durham) or "DNA AWAY" (Oxford) and rinsed in absolute ethanol.

In Durham, the samples were extracted using the protocol described in (*36*). Bone powder (100-400 mg) was digested in 0.425 M EDTA, 0.05% SDS, 0.05 M Tris-HCI and 0.333 mg/mL proteinase K and incubated overnight (18-24 hours) on a rotator at 50 °C, or until fully dissolved. The digestion buffer, excluding proteinase K, was UV-irradiated (254 nm wavelength) for an hour in a dedicated cross-linker prior to use. 2mL of extract solution was then concentrated in a Millipore Amicon Ultra-4 30 KDa MWCO (Millipore) to a final volume of 100µL. The concentrated extract was purified using silica spin-columns (QIAquick PCR Purification Kit, Qiagen) following manufacturers recommendations, except that the final elution step was performed twice to produce a final volume of 100µL. One in five or one in ten negative extraction controls were performed alongside the ancient bone samples.

In Oxford, we used 80-120 mg of bone or dentine powder that was first pre-digested in 1 mL a buffer made of EDTA and proteinase K for 1h on a rotator at 37 °C, to remove extracellular DNA and some contaminants ((*37*). After centrifugation, this first lysis was removed and stored in the freezer. The pellets were then fully digested in 1.8 mL of the same digestion buffer, incubated overnight on a rotator at 50 °C. The DNA was extracted from the resulting lysis using a modified (*38*) protocol. In brief, the 1.8 mL of solution were mixed to 50 ml of the extraction buffer 0.5 M EDTA, pH 8.0 to a final concentration of 0.45 M, 10 mg/ml Proteinase K to a final concentration of 0.25 mg/ml and Tween 20 to a final concentration of 0.05%) and centrifuged through a silica spin column (Qiagen) using an extender. Once bound to the silica filter, the DNA was purified using the commercially provided buffer (PE) and eluted in 42 µL of TET.

Library building and sequencing

The libraries were built in Oxford from the DNA extracts obtained in Durham or Oxford, without distinctions, using the Blunt End Single Tube (BEST) protocol (*39*). The amplification of the indexed DNA was optimised using quantitative PCR to estimate the best number of cycles. As expected, the number of required cycles varied largely between archaeological and museum samples. In some instances, we performed multiple parallel amplifications to increase the resulting amount of endogenous DNA. Sequencing of selected libraries was performed on a NovaSeq S4 platform (2×150 bp) at Novogene and Macrogen.

Whole genome enrichment and sequencing

Libraries from the following individuals: ASU_AnSon_1, ASU_ManBac_1, ASU_Taiwan_1, ISU_Philippines_1, ISU_Philippines_3, ISU_UchuPalau_1, ISW_Moluccas_1, POU_Vanuatu_3, POU_Vanuatu_5 were subjected to whole genome enrichment (WGE). Genomic DNA from 3 individuals: i.e. a wild boar from Japan (*Sus scrofa*), a European wild boar (*Sus scrofa*), and a bearded pig (*Sus barbatus*) from the Department of Animal Sciences, University of Illinois, was provided to Daicel Arbor Bioscience for baits construction. The gDNA was visually inspected for color and viscosity and the total DNA was quantified via a spectrofluorimetric assay. WGE baits were prepared from equal masses of the 3 gDNA samples.

Libraries were provided to Daicel Arbor Bioscience for capture. They were quantified via quantitative PCR (qPCR). Based on the quantification results (ranging from 1.9ng-445.6ng), the libraries were reamplified for 4-8 cycles to increase the mass, targeting approximately 1000ng per library. Amplified libraries were quantified again via qPCR. For libraries prepared from the same individual, equal masses of each library were pooled for capture.

Each capture pool was dried down to 7 uL by vacuum centrifugation. Captures were performed following the myBaits v5.02 protocol using the custom Pig WGE baits with an overnight hybridization and washes at 60C. Post-capture, the reactions were amplified for 10-18 cycles and were quantified with a spectrofluorimetric assay. A second round of capture was performed in the same manner. The post-capture material was visualized using the TapeStation 4200 (Agilent) platform with a High Sensitivity D1000 tape. Captures that contained adapter artifact underwent gel excision to remove the artifact. The captures were pooled in approximately equimolar ratios. Samples were screen sequenced on the Illumina NovaSeq 6000 platform on a S4 PE150 lane.

Modern samples

Papua New Guinea, Bhutan, and Sri Lankan (contact: Jaime Gongora)

DNA samples from Bhutanese (SAD_Bhutan_1-4; SAW_Bhutan_6-11), Nepalese (ASD_Nepal_1-3), and PNG (PNU_PNG_6-14) pigs were extracted from whole blood using the QIAamp® DNA Blood Mini Kit; and from hair samples using the Blood and Body Fluid Spin Protocol. DNA from Sri Lankan pigs (SAD_SirLanka_1-4; SAW_SirLanka_5) was extracted using the salting method as described by Miller et al. (*40*). DNA concentration was determined using ethidium bromide staining and a NanoPhotometer, and ranged from 25-200 ng/µl. All sampling and DNA extraction were done in the respective countries in collaboration with relevant institutions.

With the exception of seven samples (described below), libraries were prepared by the Australian Genome Research Facility (AGRF) as a commercial service using the Illumina Nextera DNA Flex library preparation protocol. Library quality was assessed using a

TapeStation and qPCR, and sequenced by AGRF on a NovaSeq 6000 S4-300 flow cell, v1.5 chemistry.

The remaining seven samples (SAW_Bhutan_6-11; SAW_SriLanka_5) were wild boars from Bhutan and Sri Lanka, collected opportunistically from dead, slaughtered, or captured individuals. DNA from these seven samples was potentially degraded because the tissue was not freshly collected. Consequently, aliquots of purified DNA extracted from these samples were sent from the University of Sydney to the Australian Centre for Ancient DNA (ACAD), University of Adelaide, where libraries were created in a clean room using a protocol optimised for fragmented, low-quality DNA. Specifically, DNA was blunt-ended and had truncated Illumina TruSeq adapters ligated following the protocol of (*41*). Full length Illumina TruSeq adapters (with a unique non-overlapping 7nt i5/i7 index combination per sample) were added via PCR using primer sequences from (*42*).

Each library was amplified in eight separate 25 µL reactions, each comprising 3 µL of undiluted library, 1 x Platinum Taq DNA Polymerase High Fidelity buffer (ThermoFisher Scientific), 2 mM MgSO4 (ThermoFisher Scientific), 0.25 mM of each dNTP (ThermoFisher Scientific), 0.4 µM of each primer, and 0.2 U of Platinum Taq DNA Polymerase High Fidelity (ThermoFisher Scientific), in laboratory grade water. Cycling conditions for the PCR were as follows: 94 °C for 2 min; 20 cycles of 94 °C for 30 s, 60 °C for 30 s, 68 °C for 40 s; and 68 °C for 10 min. PCR products from each library were pooled and purified using 1.1 x volume AxyPrep (Axygen), washed twice with 80% ethanol, and then resuspended in 30 µL of buffer comprising 10 mM Tris, 0.1 mM EDTA, and 0.05% Tween-20. Library quality was assessed using a Fragment Analyzer (Agilent) and sequenced by AGRF on a NovaSeq 6000 S1-300 flow cell, v1.5 chemistry.

Lanyu pigs (Taiwan; contact: Larry Schook)

DNA was extracted using standard phenol-chloroform techniques. Library preparation and sequencing was carried out at Macrogen (Seoul, South Korea) using an Illumina PCR-free TruSeq library kit (Illumina, Inc) followed by 2x150 bp Paired End analysis on an Illumina NovaSeq instrument.

New Caledonia (contact: Patrick Barriere) and Philippines (contact: Michael Herrera)

DNA extraction was carried out at the PalaeoBARN, University of Oxford. Three hairs per sample were cut into approximately 1 cm long sections and washed in a 0.5% sodium hypochlorite solution. Bleach was removed by rinsing the hairs twice in HPLC grade water followed by two washes in UltraPure DNase/RNase-Free Distilled Water. DNA extraction was carried out using the DNeasy Blood & Tissue Kit (Qiagen), including digestion in Buffer ATL/Proteinase K/DTT solution overnight. Single-stranded DNA sequencing libraries were prepared at the Laboratory of Functional Genome Analysis, Genzentrum, University of Munich (LMU) Swift BioSciences 1S kit. Sequencing: screening on NextSeq at Genzentrum followed by deep sequencing on a 150 PE NovaSeq 6000 S4 flowcell at Macrogen (Seoul, South Korea).

Data processing

Modern genome

Raw reads from modern genomes from both publicly available samples and newly sequenced, were first processed (i.e. adapter/PolyG tail trimmed, quality/length filtered) with fastp using default settings (see https://github.com/OpenGene/fastp) (*43*). Reads were then aligned using BWA-MEM (*44*), to the Sscrofa11.1 reference genome, duplicates were removed using Picard MarkDuplicates.

Historical and ancient genomes

Historical and ancient genomes were processed separately using the nextflow eager pipeline (*45*) with default parameters except for the following options: --mapper bwamem, and --clip_readlength 30. This includes pre-processing of reads using AdapterRemoval (*46*) to remove adapter and collapse pairs and removing PCR duplicates using MarkDuplicates. MapDamage2 (*47*) was run on the resulting BAM files to assess the level of DNA damage (Fig. S1). Most samples do not exhibit significant deamination patterns, except for the ancient DNA samples, which include the ancient Philippines (ISU_Philippines_1, and 3), Vanuatu (POU_Vanuatu_3), and Palau (ISU_UchuPalau_1) pigs. However, the degradation in the historical samples is evident from the read length distribution (Fig. S2), which shows that most reads are shorter than 200bp.

Mitochondrial analyses

Mitochondrial genome analysis

We used htsbox (https://github.com/lh3/htsbox) to create a mitochondrial genome majority consensus fasta file based on BAM files (mapped to the Ssc11.1 reference genome which contains the mitochondrial DNA as a contig), using a only reads with a mapping and base quality of 30, and at least 20 base pair long, the first and last 5 base pair were also removed for ancient and historical samples. Fasta files were merged into a single file which was filtered using a custom perl script. Sequences with less than 10x average depth of coverage, over 70% missing data were excluded from this analysis, and positions in the alignment with 30% missing data were also removed. This resulted in an alignment of 585 individuals and 15,523bp. A phylogenetic tree was then constructed with RAxML (*48*) using 100 (fast) bootstrap replicates and a GTRGAMMA substitution model (Fig. S3).

The Pacific Clade mitochondrial haplogroup was first described using short mtDNA control region sequences (*10*). To assess the monophyletic strength of this clade, we used complete mitogenomes to build a maximum likelihood phylogeny using 585 mtDNA genomes with >10x coverage. Previously described *S. scrofa* haplogroups, such as European, Near Eastern, East Asian, and a haplogroup including the mtDNA sequences from ISEA endemic *Sus* species (Fig. S3), all formed well-supported clades. We also identified two well-supported clades corresponding to the Pacific Clade and the Lanyu Clade, confirming their identification as monophyletic groups using short-fragment PCR data (*10*).

Published mtDNA control region dataset

Published mitochondrial control region sequences, many of which belong to the same individuals for which we obtained nuclear data in this study, (N =707, see Supplementary Data S2) were aligned using global iterative refinement (G-INS-i) implemented in MAFFT 7.526 (2024/Apr) (*49*). The alignment was visualised and a 621-bp segment of the control region was selected using Geneious Prime® 2024.0.7 (GraphPad LLC). IQ-TREE multicore version 1.6.12 (*50*) was used to construct a maximum likelihood tree under a HKY+F+R4 model of evolution, selected using ModelFit (*51*). Branch support was assessed using 1000 Ultrafast bootstrap psuedo-replicates (*52*). The bootstrap consensus tree was drawn in FigTree 1.4.5 (https://github.com/rambaut/figtree) and used to assign sequences to one of four haplogroups: Lanyu, Pacific Clade, Europe (European *S. scrofa* in Fig. S3), East Asia (Asian *S. scrofa* in Fig. S3) , and ISEA (ISEA-endemic *Sus*, i.e. *S. barbatus, S. philipensis/S. cebifrons, S. verrrucosus,* or *S. celebensis* in Fig. S3).

Nuclear analyses

Reference panel genotype call

Genotyping was performed using graphtyper v2.7.2 (*53*) using genomes with at least 5x (500 individuals). The resulting VCF was filtered using the recommended following options (*53*):

vcffilter -f "ABHet < 0 | ABHet > 0.33 & ABHom < 0 | ABHom > 0.97 & MQ > 30"
(requiring allelic balance of 33% for heterozygous and >97% for homozygous and a mapping quality of 30)

bcftools view -f "PASS" -m 2 -M 2 -v snps
(selecting on biallelic SNPs flagged as PASS)

bcftools-1.11/bcftools filter --SnpGap 10 -i "FORMAT/GQ>30" -i "INFO/AC>4"
(removing every SNPs within 10bp of an indel, and requiring genotyping quality of 30 and allele count of at over 4 reads).

This resulted in a total of ~95M SNPs, including ~25M transversions.

Phasing and imputation

We used SHAPEIT v5 for phasing (*27*) the reference panel which contains 95M SNPs in 500 individuals. Phasing was performed in 1Mbp sliding windows (100Kb overlap) using the phase_common tool and a recombination map obtained from (*54*). The resulting BCF were ligated using the ligate tool.

We then used GLIMPSE 2 for imputation (*28*). We first used the GLIMPSE2_chunk tool to split the phased phase into imputation regions based on the recombination map obtained from (*54*). Each chunk was then converted into a binary reference format using the GLIMPSE2_split_reference tool. Each BAM file for samples between 1-5x was then imputed separately using GLIMPSE2_phase for each chunk, and chunks were ligated using GLIMPSE2_ligate. The resulting BCFs were then merged with the panel using bcftools. Imputation accuracy was estimated by first downsampling the BAM files of nine individuals to 0.1x, 0.5x, 1x, 2x and 4x using samtools, which were selected to represent the diverse ancestry in our dataset i.e.: ASD_Luchuan_3 and ISD_Lanyu_1 (Asian S. scrofa), EUD_Pietrain_4 (European domestic), CEL_Sus−Celebensis_1 (ISEA-endemic Sus), INDO_Sumatra_1 and ISW_Bali_1 (Indonesia S. scrofa/ISEA-endemic *Sus* hybrid), ISW_Buru_1 and PNU_PNG_10 (Wallacean *S. scrofa*) and POW_NewCaledonia_4 (hybrid Asian and European *S. scrofa*).

We then imputed each bams using a reference panel from which these six individuals were removed. We then calculated concordance between “true” genotypes and imputed genotypes (r2) at different minor allele frequency (MAF) using the GLIMPSE_concordance tool (Fig. S4). Imputation accuracy was >95% in most individuals from at 2x and MAF>1%. Island Southeast Asian feral individuals from Bali (ISW_Bali_1, ISW_Buru_1), however, only achieved an accuracy of >90% at MAF>5%.

The primary objective of imputation in this analysis is to conduct ancestry deconvolution through local ancestry inference. In addition to testing accuracy of imputation on genotypes, we investigated how imputation influenced our ability to conduct local ancestry inference (LAI) using GNOMIX, to pinpoint genomic segments associated with specific ancestries (see below for details on GNOMIX analysis). We utilized the imputed VCFs derived from downsampled BAM files in the preceding step to conduct a LAI using GNOMIX. For each ancestry, we then compared the regions identified by gnomix in both the full coverage and downsampled data. The results revealed a high degree of overlap between

these regions, ranging from 94% to 99% (Fig. S5). This suggests that imputation minimally impacted the power of local ancestry inference in this context.

Pseudohaploid call

The 25M transversion SNPs discovered in the 5x data set were then genotyped in all historical and ancient genomes (i.e. those below 5x depth of coverage) not included in the panel using ANGSD (v0.933) with the following options: -dohaplocall 1 -minMinor -doCounts 1 -minMapQ 20 -minQ 20 -minInd 1 -setMinDepth 1 -trim 5. Positions at which the genotype did not match one of the two alleles in the 25M SNP panel were set as missing data. Genotype data were merged using plink v1.9 (*55*). Samples from the panel were pseudohaplosised using a custom script in the resulting plink file.

Principal component analysis (PCA)

Principal component analysis (PCA) was carried out using emu (*56*) on pseudohaploid transversions. We filtered SNPs with MAF<5% and kept all samples with >10,000 SNPs covered (576 individuals).

ADMIXTURE

We first ran an unsupervised ADMIXTURE clustering based on plink file containing data from 120 pseudohaploid genomes from wild individuals form East Asia, Europe and ISEA to identify individual carrying >99% ancestry from three different clusters: Asian *S. scrofa*, European *S. scrofa* and ISEA-endemic *Sus* (Fig. S6). Individuals with at least 99% ancestry from one cluster were then used as a source in a supervised ADMIXTURE analysis (K=3 again), based on 576 individuals.

We also ran an unsupervised ADMIXTURE clustering based on 304 pseudohaploid genomes. All genomes were included in this analysis, except European domestic pigs and no more than 3 individuals per Asian breed. We used European wild as a proxy for European ancestry rather than European domestic pigs because many of the latter possess Asian ancestry (*57*) At K=3 (best K based on cross-validation), the results (Fig. S7) mirrored those in Fig. S6 (only using wild boar), showing the impact of European ancestry (blue) in Asian domestic (ASD) and some individuals East of the Wallace Line (see also the blue component in Fig. 2), and the impact of ISEA-endemic *Sus* ancestry (black) in ISEA individuals from both East and West of the Wallace Line (see also Fig. 2).

At K=5 (second best K based on cross-validation) this analysis separated ISEA individuals East (mostly possessing the red ancestry component) and West of the Wallace Line (mostly possessing the light green ancestry component).

D and F statistics

Our ADMIXTURE analysis with 120 wild pig genome did not possess enough power to split the ancestry of ISEA-endemic *Sus* individual i.e. *S. celebensis, S. cebifrons, S. verrucosus,* or *S. barbatus*. Reducing the number of *S. scrofa* individuals in ADMIXTURE to 25 allowed for the separation of each species ancestry at K=9 (Fig. S8) demonstrating that this lack of power in our full analysis is likely to be due to an ascertainment bias due to the comparatively smaller number of ISEA-endemic *Sus* genomes in our dataset (n=23).

Therefore, for each ISEA individual that possessed both Asian *S. scrofa* and ISEA-endemic *Sus* based on our ADMIXTURE analysis we first used D-statistic to assess which ISEA-endemic *Sus* best matched their non-*scrofa* ancestry. To do so we first validated that each individual yielded significant D-statistics of the form D(Babyrousa outgroup, ISEA-endemic Sus; Test, Asian *S. scrofa*), where ISEA-endemic *Sus* represent any genome of *S.*

celebensis, *S. cebifrons*, *S. verrucosus*, or *S. barbatus* and Asian *S. scrofa* any genome of individuals that possess >99% Asian *S. scrofa* ancestry in our ADMIXTURE analysis.

We then ran all combinations of D-statistics of the form D(B. babyrousa, Test; ISEA-endemic-Sus-Y, ISEA-endemic-Sus-Z), where Test is the individual possessing *S. scrofa* and ISEA-endemic *Sus* ancestry, and ISEA-endemic-Sus-Y, and ISEA-endemic-Sus-Z represent all combinations of ISEA-endemic *Sus* individual i.e. *S. celebensis, S. cebifrons, S. verrucosus,* or *S. barbatus*. We choose the individual species with the highest D value as the source of ISEA-endemic *Sus* ancestry ancestry which was used for the plot in Fig. 2.

Admixture graphs

We suspected that Western Indonesian mixed *S. scrofa* / ISEA-endemic *Sus* (e.g. from Sumatra and nearby islands) possess *S. scrofa* ancestry that is poorly represented by our panel. To address this, we constructed admixture graphs using Admixtools2 (https://github.com/uqrmaie1/admixtools), to assess the phylogenetic placement whether the *S. scrofa* ancestry found in these individuals is nested within Asian *S. scrofa* or a sister group to Asian and European *S. scrofa* (Fig. S9).

We constructed "skeleton graphs", tailored to for each *S. scrofa* ISEA genome, informed by the D-statistic pipeline above. For individuals with no detectable I ISEA endemic-Sus ancestry, we constructed admixture graphs containing a *B. babyrousa* as an outgroup, an individual with Asian *S. scrofa* ancestry, an individual with European *S. scorfa* ancestry, and an individual with South Asian *S. scrofa* ancestry (i.e. Sri Lanka or India). The graph was created using the find_graphs function (plusminus_generations = 5, stop_gen = 50, stop_gen2 = 10). We consistently identified the same topology: (Babyrousa, (South Asian, (Asian, European)))

For individuals with detectable ISEA endemic-Sus ancestry we carried out the same as above, but included a genome corresponding to the relevant non-scrofa species (i.e. the species that was the best proxy for the ISEA endemic-Sus ancestry based on D-statistics). This also consistently identified the same topology: (Babyrousa, (ISEA endemic-Sus, (South Asian, (Asian, European))))

We then added each test genome (i.e. *S. scrofa* genome from ISEA) to its corresponding skeleton graph from 1a or 1b, above, in two different locations to create two admixture graphs for each test genome:

- basic graph 1a/1b-Asian: On the Asian branch, i.e. (Babyrousa, (South Asian, ((Asian, test), European))), 1a; or (Babyrousa, (ISEA endemic-Sus, (South Asian, ((Asian, test), European)))), 1b
- basic graph 1a/1b-Sister: On the basal branch, i.e. (Babyrousa, (ISEA endemic-Sus, (South Asian, (test, (Asian, European))))) or (Babyrousa, (South Asian, (ISEA endemic-Sus, (Asian, European))))

For the basic graphs for individuals with detectable non-*scrofa* ancestry (basic graph 1b-Asian & 1b-Sister), we added an admixture edge between ISEA endemic-*Sus* and our test genome.

For each individual (test), we then assessed whether the Asian or Sister graph was the best fit, using 100 bootstrap-resampled graph fits, after computing out of sample scores to fairly compare models of different complexity to each other.

The ratio of scores for the two admixture graphs were plotted in Fig. S9, as:
(Score_Asian - Score_Sister) / Max[Score_Asian, Score_Sister]

Local ancestry reconstruction (GNOMIX)

Local ancestry reconstruction was performed using GNOMIX (*29*). Wild individuals with Asian *S. scrofa*, European *S. scrofa* and non-*scrofa* ancestry that were used as a source in the supervised ADMIXTURE above were used as reference for this analysis.

We trained two models, each with two sources:

- *S. scrofa* (using European *S. scrofa*) and non-*scrofa*
- European *S. scrofa* and Asian *S. scrofa*.

For model 1, we did not use Asian *S. scrofa* to discriminate between *scrofa* and non-*scrofa* ancestry as previous studies have suggested that mainland Southeast Asian wild boar (*S. scrofa*) possess some small degree of ISEA-endemic *Sus* ancestry (*3*, *4*).

We ran GNOMIX with default settings except for the time of admixture used for validation (generation since admixture: 2, 4, 6, 8, 12, 16, 24, 48). Both models achieved training accuracy and validation accuracy > 99%. Local ancestry reconstruction was then performed on *S. scrofa* individuals from Asia that were either phased (>5x) or imputed (>1x) and which were showed to possess either European *S. scrofa* or non-*scrofa* ancestry (non possessed both), based on the ADMIXTURE analysis, using the appropriate model (model 1 for non-*scrofa* ancestry, model 2 in case of European *S. scrofa* ancestry). See the section on imputation above for details about the results of the test for accuracy of imputation for LAI.

Ancestry deconvolution

We used the GNOMIX results to identify regions of Asian *S. scrofa* ancestry in the genome of *S. scrofa* individuals with mixed ancestry, i.e. possessing either non-*scrofa* and Asian *S. scrofa* (GNOMIX model 1) or European *S. scrofa* and Asian *scrofa* ancestry (GNOMIX model 2). To do so, we selected regions which were identified with high confidence (probability >99%) by GNOMIX as either homozygous *S. scrofa* (model 1) or homozygous Asian *S. scrofa* (model 2). For each mixed individual, we extracted these regions from the pseudohaploid PLINK file used for PCA, ADMIXTURE, and D-statistics analyses described above.

We then calculated admixture fraction using f4 ratio to assess the proportion of non Asian *S. scrofa* ancestry left in the genomes of mixed individuals after deconvolution (Fig. F4-deconvolution). European ancestry in most (14/16) individuals from New Caledonia, Hawai, and Papua New Guinea decreased from ~50-25% to 0-5%. European ancestry in one individual from New Caledonia, however, remained at ~10%, at ~20% in an individual from Papua New Guinea. European ancestry in Lanyu pigs decreased from ~25% to 0-1%, while European ancestry in domestic pigs from the Philippines decreased from 30-70% to 0-15%.

Non-*scrofa* ancestry in individuals with lower *S. scrofa* ancestry such as pigs from Sibago or pigs East of the Wallace line, including Flores, Timor, and Halmhaera decreased from between ~80-40% to below 5%. The non-*scrofa* ancestry in wild boars from Western Indonesia (West of the Wallace line), including Sumatra and nearby islands, however, decreased from between 14-5% to ~5%. The lower success of deconvolution in these individuals can likely be explained by ancient admixture processes (18) which would be found in smaller blocks of non-*scrofa* ancestry across the genome that are undetectable by GNOMIX. In contrast, due to its recent introduction in the region, European *S. scrofa* ancestry is found in larger blocks (Fig. 3B), which are more easily detectable.

We also conducted a PCA using both full genome and deconvoluted data using all *S. scrofa* individuals from East Asia (n=228). Fig. S10A, which includes all Asian *S. scrofa* individuals without ancestry deconvolution, reveals a separation between populations east and west of the Wallace Line. Some individuals from the East, however, such as individuals

from Timor and Halmhaera, cluster closer to pigs from west of the Wallace line likely reflecting the varying degrees of ISEA-endemic ancestry. These individuals instead plot close to pigs East of the Wallace Line following ancestry deconvolution (Fig. S10B).

This Fig. highlights that *S. scrofa* populations on either side of this biogeographic boundary possess markedly different ancestry profiles, despite their relative geographic proximity. Populations east of the Wallace Line show a closer affinity to mainland domestic populations, particularly those from Southeast China such as Luchuan, Wuzhishan, and Bamaxiang.

outgroup f3

We calculated outgroup-f3 statistics using calc-f3 from the structf4 package (*58*) and babirusa as an outgroup. The goal of this analysis is to assess the degree to which Lapita pigs from Melanesia share drift with other populations of *S. scrofa*. To do so we first chose a 2,500years old ancient Vanuatu genome (POU_Vanuatu_3; ~0.7x) as reference and computed all pairwise outgroup-f3 involving this individual.

We found that feral pigs from Papua New Guinea (e.g. ISW_PNG_18; ~2.2x) shared the most drift with this individual (Fig. S11). ADMIXTURE analysis of ISW_PNG_18 did not detect any evidence of European ancestry, making it a good proxy for Lapita pig ancestry. We used this individual as reference in subsequent outgroup-f3 analyses due to its higher coverage than POU_Vanuatu_3 to limit issues arising when comparing multiple individuals with low breadth of coverage, particularly after ancestry deconvolution. We then computed all pairwise outgroup-f3 involving ISW_PNG_18 using both the full and deconvoluted dataset.

Admixture time (DATES)

We used DATES (*26*) to estimate time of admixture in admixed Island Southeast Asian individuals using the recombination map obtained from (*54*) and the following parameters:

binsize: 0.001
maxdis: 1
jackknife: YES
qbin: 10
runfit: YES
afffit: YES
lovalfit: 0.45

The same individuals used to train the GNOMIX models were used as source for this analysis. We used a generation time of 3 years to convert into years - admixture date were subtracted to age of samples (specific age for historical genomes in Data S1 and using year 2000 for modern samples). Results were considered significant at first if Z-score > 2 and normalized root-mean-square deviation (NRMSD) < 0.7. We first assessed the accuracy of DATES using five European domestic individuals with known dates of admixture with Asian *S. scrofa* (as a result of 19th century improvement of pig stock in Europe). We modelled these individuals using the same source as for the ADMIXTURE and GNOMIX analyses.

All results were significant and visual inspection of the exponential distribution fitted and the covariance decay curve suggested a good fit (Fig. S12). The recovered admixture proportion and admixture dates admixture dates fitted well with the known history of these breeds (admixture date in the middle of the 19th century (*59*)):

Sample ID	Sources	EUW	ASW	Mean time (generation)	SE	Z	NRMSD	Mean time (years)
EUD_Duroc_1	EUW-ASW	0.78	0.22	55.5	15.0	3.7	0.083	1833
EUD_Landrace_11	EUW-ASW	0.784	0.216	39.5	8.5	4.6	0.058	1881
EUD_Pietrain_10	EUW-ASW	0.772	0.228	46.8	8.1	5.8	0.069	1860
EUD_Tamworth_3	EUW-ASW	0.831	0.169	59.9	14.4	4.1	0.07	1820
EUD_Yorkshire_19	EUW-ASW	0.758	0.242	53.7	7.8	6.9	0.073	1839

We modelled individual from Oceania, Melanesian and the Philippines using the same approach:

Sample ID	Sources	EUW	ASW	Mean time (generation)	SE	Z	NRMSD	Mean time (years)
ISD_Lanyu_1	EUW-ASW	0.351	0.649	6.221	1.987	3.131	0.137	1981
ISD_Lanyu_2	EUW-ASW	0.29	0.71	8.63	7.832	1.102	0.077	1974
ISD_Lanyu_3	EUW-ASW	0.319	0.681	3.147	6.363	0.495	0.094	1991
ISD_Lanyu_4	EUW-ASW	0.342	0.658	-2.435	10.307	-0.236	0.136	2007
ISD_Lanyu_5	EUW-ASW	0.308	0.692	-5.063	13.104	-0.386	0.115	2015
ISU_Philippines_10	EUW-ASW	0.341	0.659	18.075	2.469	7.321	0.087	1946
ISU_Philippines_11	EUW-ASW	0.304	0.696	38.298	15.299	2.503	0.107	1885
ISU_Philippines_12	EUW-ASW	0.647	0.353	16.761	11.669	1.436	0.073	1950
ISU_Philippines_13	EUW-ASW	0.739	0.261	24.639	3.955	6.229	0.091	1926
ISU_Philippines_14	EUW-ASW	0.696	0.304	36.515	11.697	3.122	0.088	1890
ISU_Philippines_15	EUW-ASW	0.753	0.247	18.08	9.237	1.957	0.078	1946
ISU_Philippines_4	EUW-ASW	0.673	0.327	6.702	5.378	1.246	0.166	1980
ISU_Philippines_5	EUW-ASW	0.54	0.46	18.577	8.305	2.237	0.077	1944
ISU_Philippines_6	EUW-ASW	0.739	0.261	10.952	16.668	0.657	0.109	1967
ISU_Philippines_7	EUW-ASW	0.648	0.352	40.351	11.847	3.406	0.087	1879
ISU_Philippines_8	EUW-ASW	0.475	0.525	36.789	9.636	3.818	0.067	1890
ISU_Philippines_9	EUW-ASW	0.791	0.209	63.41	11.33	5.597	0.05	1810
ISW_Marquesas_1	EUW-ASW	0.639	0.361	34.074	9.877	3.45	0.096	1898
ISW_Morotai_1	EUW-ASW	0.177	0.823	30.923	6.229	4.964	0.074	1810
ISW_Morotai_2	EUW-ASW	0.186	0.814	30.347	6.181	4.91	0.072	1812
ISW_PNG_20	EUW-ASW	0.208	0.792	16.317	5.572	2.929	0.091	1922

ISW_Sumba_1	EUW-ASW	0.453	0.547	17.65	2.946	5.991	0.189	1877
PNU_PNG_10	EUW-ASW	0.424	0.576	26.165	4.892	5.349	0.058	1922
PNU_PNG_11	EUW-ASW	0.421	0.579	7.609	8.702	0.874	0.108	1977
PNU_PNG_12	EUW-ASW	0.544	0.456	25.227	3.32	7.599	0.117	1924
PNU_PNG_13	EUW-ASW	0.571	0.429	25.036	8.263	3.03	0.076	1925
PNU_PNG_14	EUW-ASW	0.482	0.518	17.289	8.536	2.025	0.078	1948
PNU_PNG_6	EUW-ASW	0.487	0.513	6.311	17.201	0.367	0.101	1981
PNU_PNG_7	EUW-ASW	0.334	0.666	21.447	10.853	1.976	0.077	1936
PNU_PNG_8	EUW-ASW	0.418	0.582	30.253	13.904	2.176	0.073	1909
PNU_PNG_9	EUW-ASW	0.488	0.512	14.085	4.373	3.221	0.13	1958
POW_Hawaii_1	EUW-ASW	0.585	0.415	40.041	8.144	4.916	0.078	1817
POW_NewCaledonia_1	EUW-ASW	0.57	0.43	23.775	4.149	5.73	0.081	1942
POW_NewCaledonia_2	EUW-ASW	0.462	0.538	33.71	7.245	4.653	0.081	1912
POW_NewCaledonia_3	EUW-ASW	0.544	0.456	41.317	8.477	4.874	0.071	1889
POW_NewCaledonia_4	EUW-ASW	0.544	0.456	30.252	9.299	3.253	0.082	1922
POW_NewCaledonia_5	EUW-ASW	0.587	0.413	23.563	15.314	1.539	0.08	1942
POW_NewCaledonia_6	EUW-ASW	0.45	0.55	31.695	7.888	4.018	0.045	1918

Most significant (Z>2) results points to dates in from the early 18th century to the mid 20th century. We attempted to estimate time of admixture between *S. scrofa* and endemic-Sus in Western Indonesia populations but visual inspection of the exponential distribution fitted and the covariance decay curve suggested a poor fit (Fig. S13). This is likely due to the fact that there are only the low number of individual genomes available for each ISEA-Endemic *Sus* species and possibly that the Asian *S. scrofa* population used here is a poor proxy for the ancestry in these populations;

Conditional heterozygosity

We calculated conditional heterozygosity (*60*) to obtain estimates of heterozygosity. This approach offers the advantage of enabling comparisons across datasets with varying sequencing coverage. Conditional heterozygosity can also be computed post-ancestry deconvolution. To do so, we ascertained heterozygous sites in a ~29x *S. cebifrons* genome (CEB_Sus-Cebifrons_10) to identify 554,543 heterozygous transversions. We chose this species as it is an outgroup to *S. scrofa* and not involved in admixture with *S. scrofa*. This strategy effectively mitigates potential biases arising from the variable levels of sequencing coverage inherent in our ancient, historical, and modern data.

For each *S. scrofa* (or hybrid) individual with >1x depth of coverage in our dataset, we randomly sampled two reads per site and computed the number of sites at which each alleles differed, excluding sites which possess alleles not present in the *S. cebifrons* individual. We recomputed this metric using homozygous tracks of Asian *S. scrofa* ancestry (i.e. masking regions of the genome with European *S. scrofa* or ISEA-endemic *Sus* identified by GNOMIX).

The result shows the influence of ancestry deconvolution, which substantially reduces heterozygosity levels (Fig. S14). This effect is particularly pronounced in individuals with ancestry derived from both Asian *S. scrofa* and ISEA-endemic *Sus* species, such as those from the islands of Timor (e.g. ISW_Timor_3 and ISW_Timor_8) or Halmahera (ISW_Halmahera_1). These individuals exhibit markedly inflated heterozygosity, which was effectively normalized when excluding their ISEA-endemic *Sus* ancestry, bringing their heterozygosity levels within the typical range of Eastern ISEA populations.

This analysis shows that with the exception of the insular domestic Lanyu pigs (a small population restricted to Taiwan), populations east of the Wallace Line possess lower heterozygosity compared to other Asian populations, including mainland Asian domestic pigs, from which they originated (wilcoxon test, $W=2332$, $p<0.001$).

Geometric Morphometric (GMM)

Data acquisition

Information on the specimens analysed with geometric morphometrics can be found in Data S3. This study used geometric morphometrics to analyze the shape and size of third lower molars. Digital images of the teeth were taken from above (occlusal view) using a standardized protocol (*61*). Twelve anatomical landmarks and 87 sliding semi-landmarks (Fig. S15) were then placed on these images using TPSdig software (*62*). Coordinates were superimposed using a General Procrustes Analysis (function ProcSym of the R package Morpho (*63*)). During this procedure the position of the SSL were adjusted using the Procrustes distance (bending = FALSE) and the original configuration was used in all iterations (recursive=FALSE). The centroid size and Procrustes residuals (coordinates after superimposition) were used respectively to assess tooth size and variation.

Predictive Discriminant Analysis

In order to identify the ancient teeth corresponding to *S. scrofa* several regional predictive linear discriminant analyses were performed. The identification was based on a resampling approach based on groups of the same size combined with a data reduction procedure selecting the first components (PC) of a Principal Component Analysis (PCA) that maximise the between group differentiation (*64*). Among the 100 analyses computed, only the discriminant analyses with the highest correct cross-validation percentages (CVP, jackknife) were retained (upper quartile of the distribution). The specimens of *S. scrofa* included in those identifications originated from China, India, Thailand, Vietnam, Malaysia, Burma, Sumatra and Java. We then classified individual samples from Island Southeast Asia using this approach.

In Sulawesi, the 4 ancient specimens could correspond either to *S. scrofa* (N=97) or *S. celebensis* (N=93). The two modern species have a correct cross-validation of 96% (90% confidence interval CI: 95.7%-96.8%) using the 17 first PCs carrying 91.3% of the total variance. For the prediction, 67 discriminant analyses were retained all with a CVP above 96%. The four specimens were identified as *Sus scrofa*.

In Sarawak (Borneo), 26 archaeological specimens were compared to 67 *S. barbatus* and 97 *S. scrofa*. The CVP of the two species is 90.7% (CI: 88.7-93.2) based on 12 PCS carrying 85.9% of the total variance. 39 discriminant analyses were retained, all with a CVP above 91%. 13 archaeological specimens were identified as *S. barbatus* and 13 to *S. scrofa*.

For the Philippines, the 48 archaeological specimens were identified as either *S. scrofa* (N=97) or ISEA endemic *Sus* (N=38, including *S. cebifrons* (N=8), *S. oliveri* (N=3), and *S. philippensis* (N=27)). The two modern taxa have a CVP of 92.7% (CI: 88.1-97.4%) based on 13 PCs (% of total variance). 33 discriminant analyses were retained resulting in the identification of 35 *S. scrofa* and 13 endemic Sus.

In Flores, 6 archaeological specimens were identified as either *S. celebensis* (N=93) or *S. scrofa* (N=97). The cross validation between the two species was 96.3% (CI: 96.2, 96.8%) based on 13 PCs (% of variance). 97 discriminant analyses were retained, with a CVP above 96.2%, that allowed to identify 5 *S. celebensis* and 1 *S. scrofa*.

The 147 archaeological specimens from mainland South-East Asia (China, Vietnam, Thailand) strongly differ from the 31 from the far east of the distribution (Marqueses, Vanuatu). Between those two groups the correct cross validation is 97% (CI: 92.3-100%) based on the first 10 PCs (83% of total variance). The Wallacean and Pacific specimens exhibited a third lower molar proportionally wider and shorter compared to the more rectangular shape of the specimens from mainland Asia (Fig. S16). Based on those PCs and the predictive discriminant analysis approach described above (see (*64*)) the remaining 102 specimens were identified based on 42 discriminant analyses with a CVP above 98%.

The morphometric proximity between ancient and modern groups from the various regions and taxonomy was visualised using Mahalanobis distances obtained from the 15 first PC scores (88.7% of total variance) that maximise the between group differences (Fig. 1C).

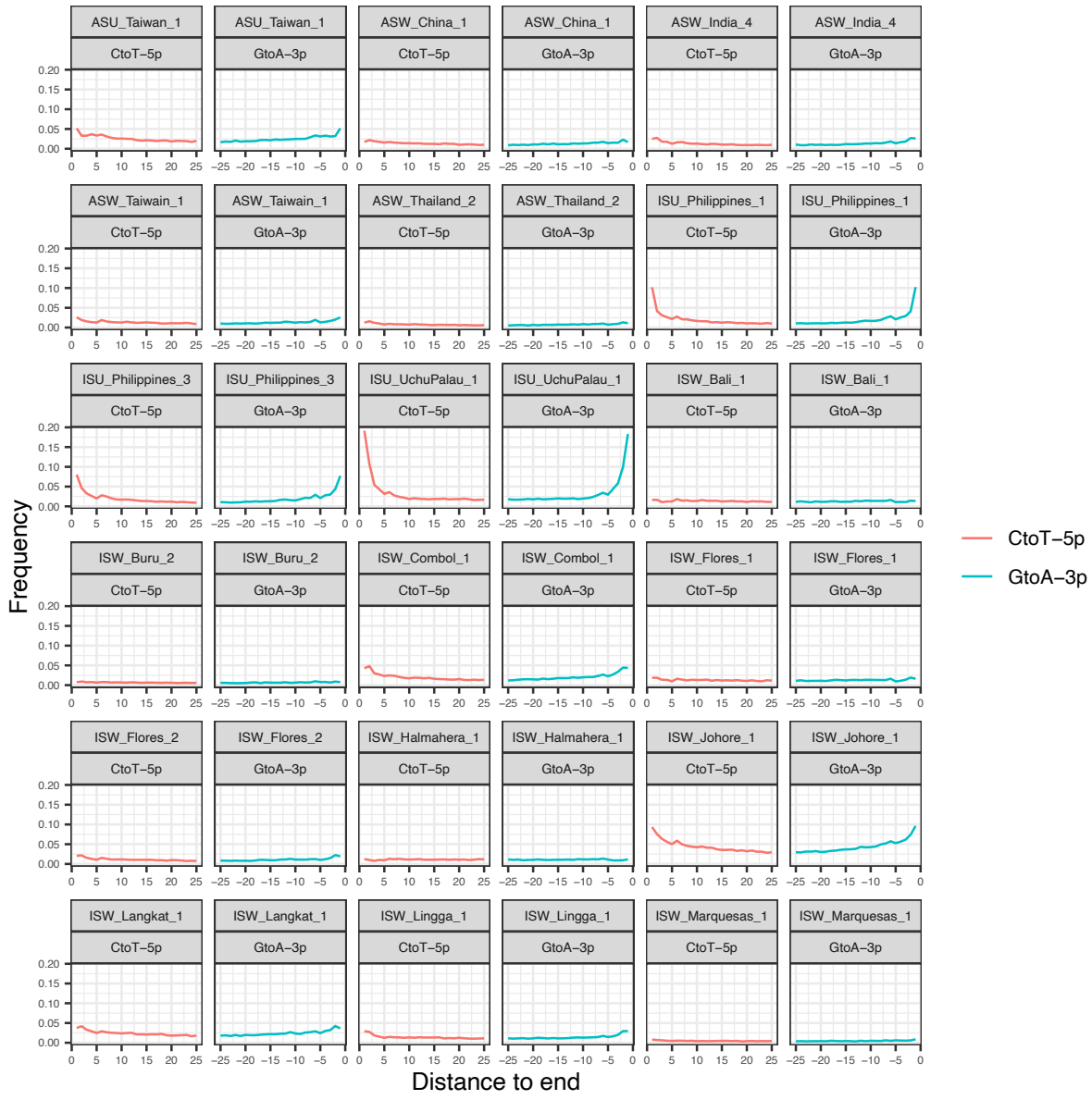


Fig. S1

C-to-T (left, red) and G-to-A (right, blue) substitution frequencies estimated for ancient and historical samples obtained from MapDamage2.

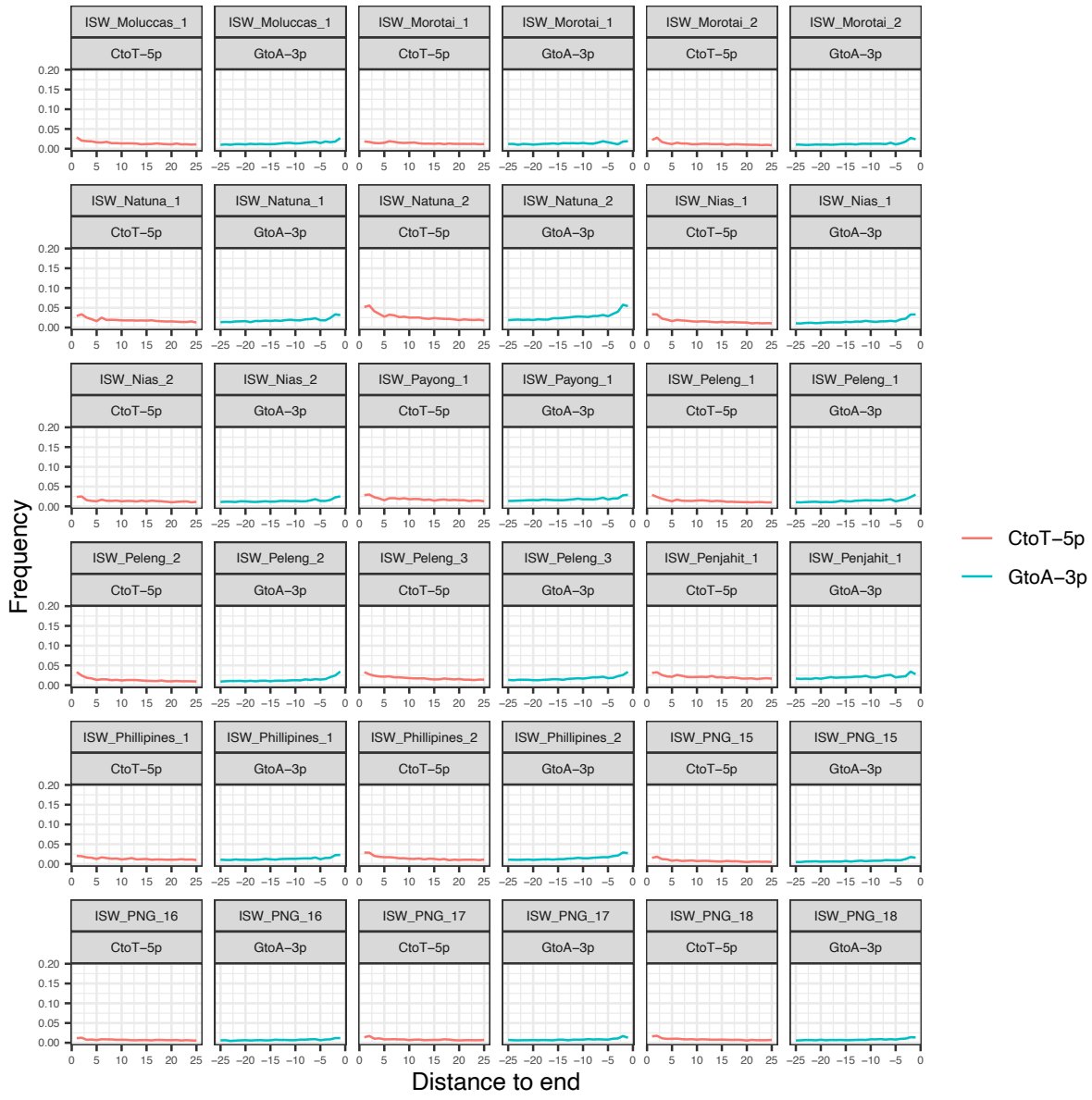


Fig. S1 (continued)

C-to-T (left, red) and G-to-A (right, blue) substitution frequencies estimated for ancient and historical samples obtained from MapDamage2.

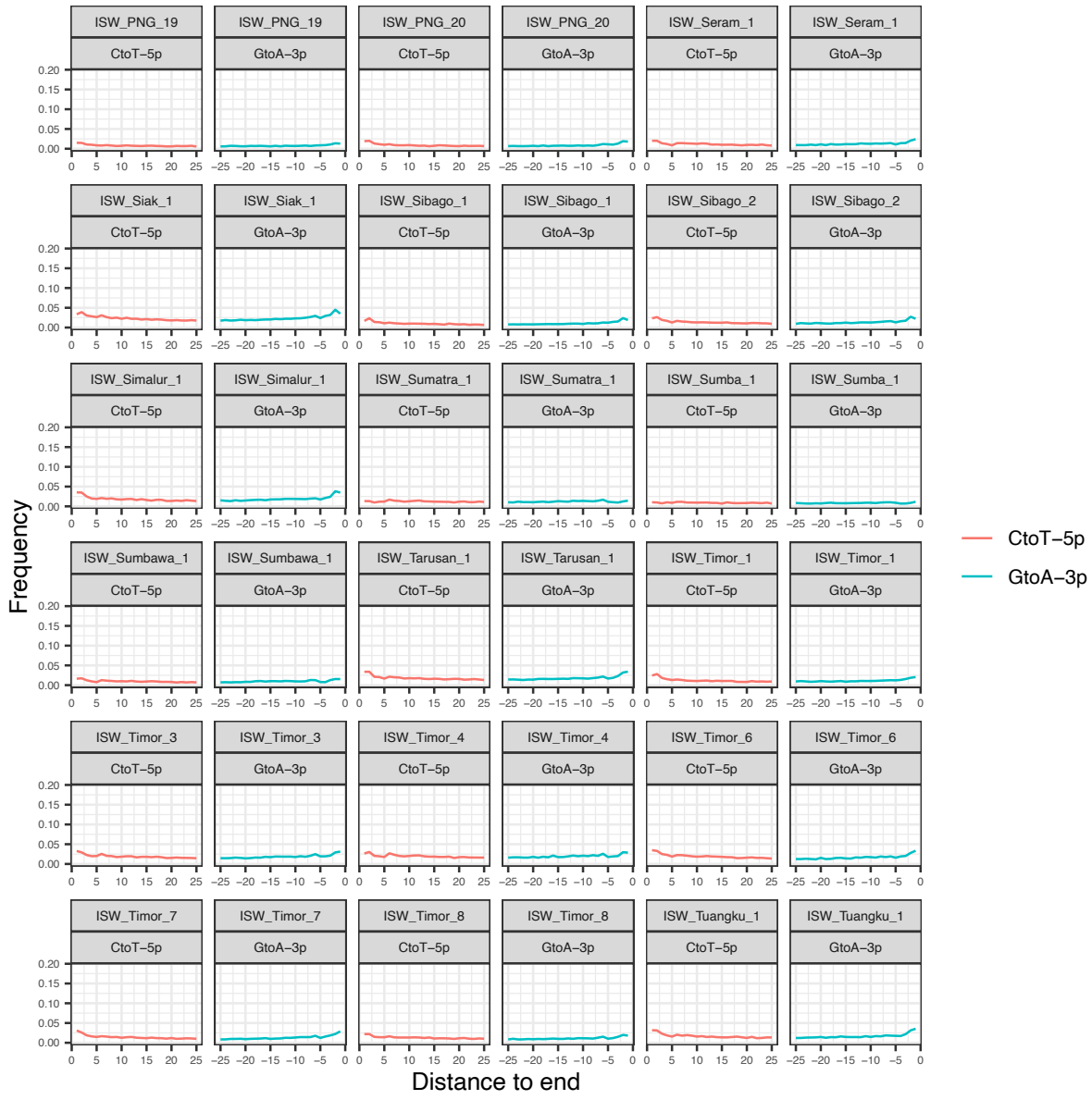


Fig. S1 (continued)

C-to-T (left, red) and G-to-A (right, blue) substitution frequencies estimated for ancient and historical samples obtained from MapDamage

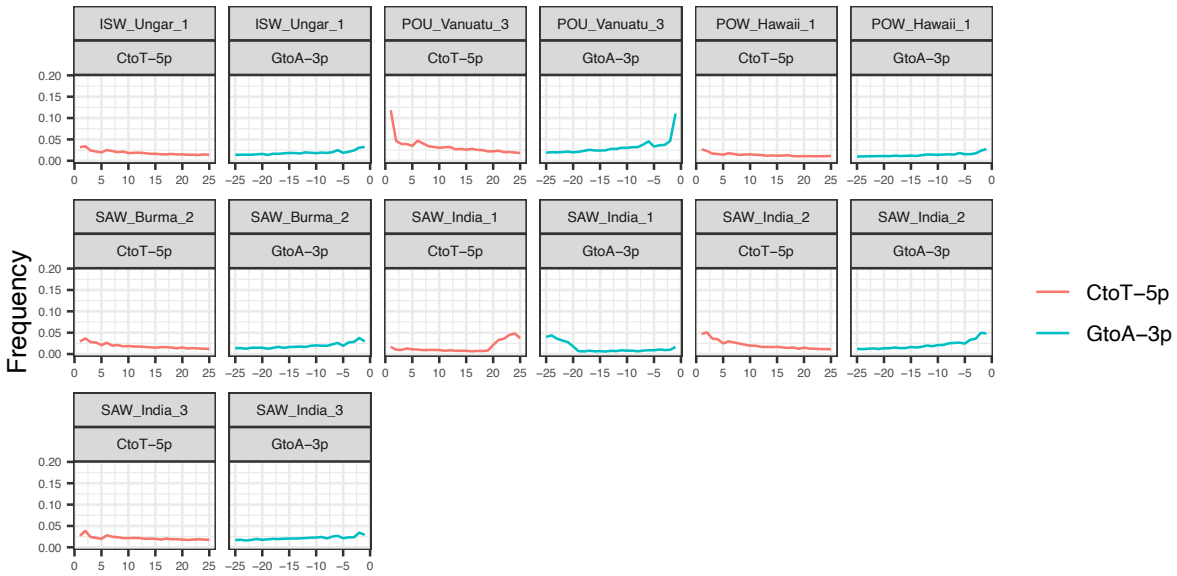


Fig. S1 (continued)

C-to-T (left, red) and G-to-A (right, blue) substitution frequencies estimated for ancient and historical samples obtained from MapDamage2

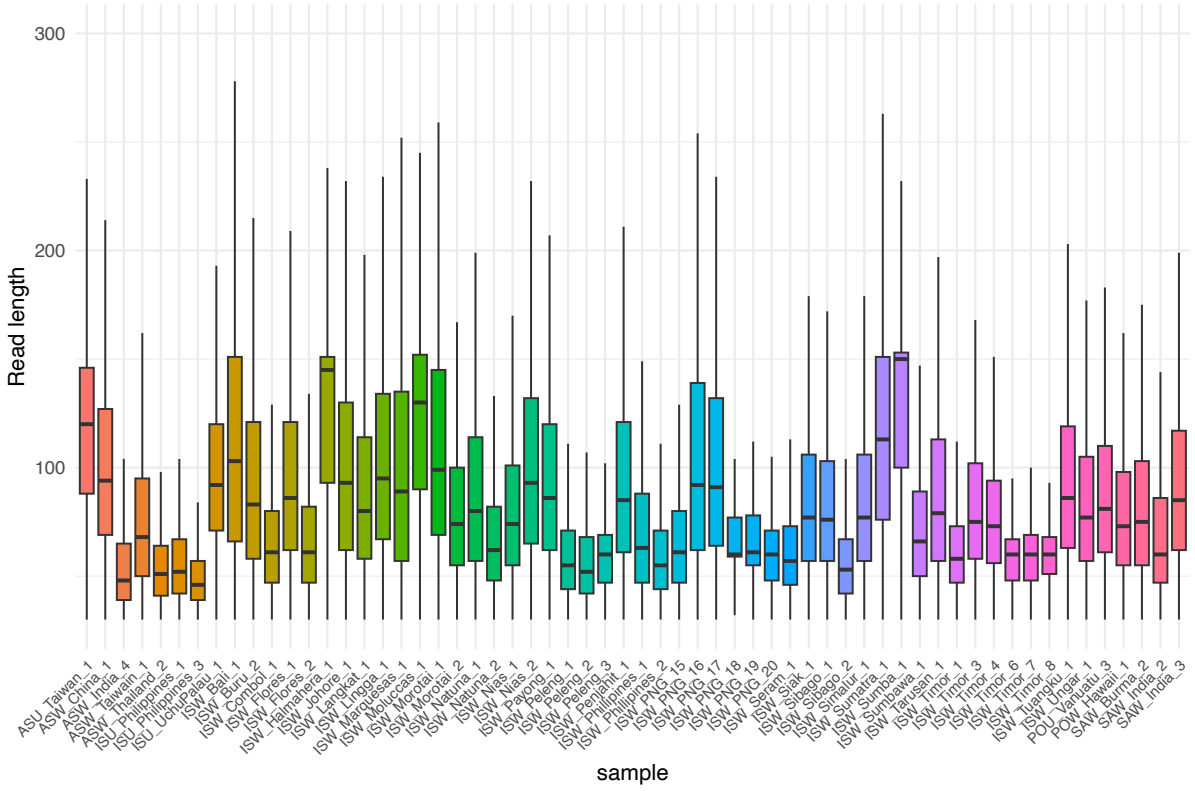


Fig. S2.
Distribution of mapped read length for each ancient and historical sample (x-axis).

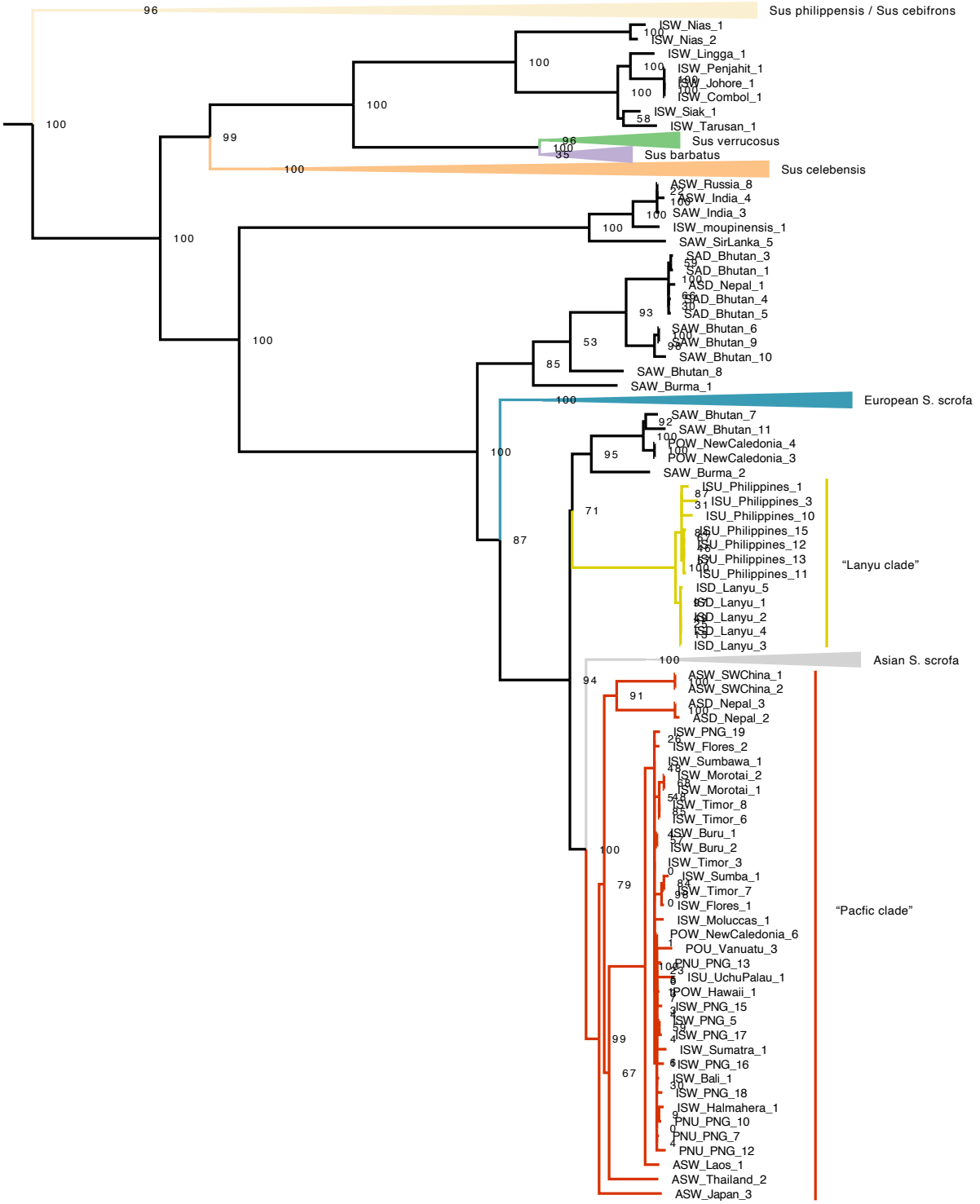


Fig. S3.
Maximum likelihood tree based on 585 mitogenomes with bootstrap support.

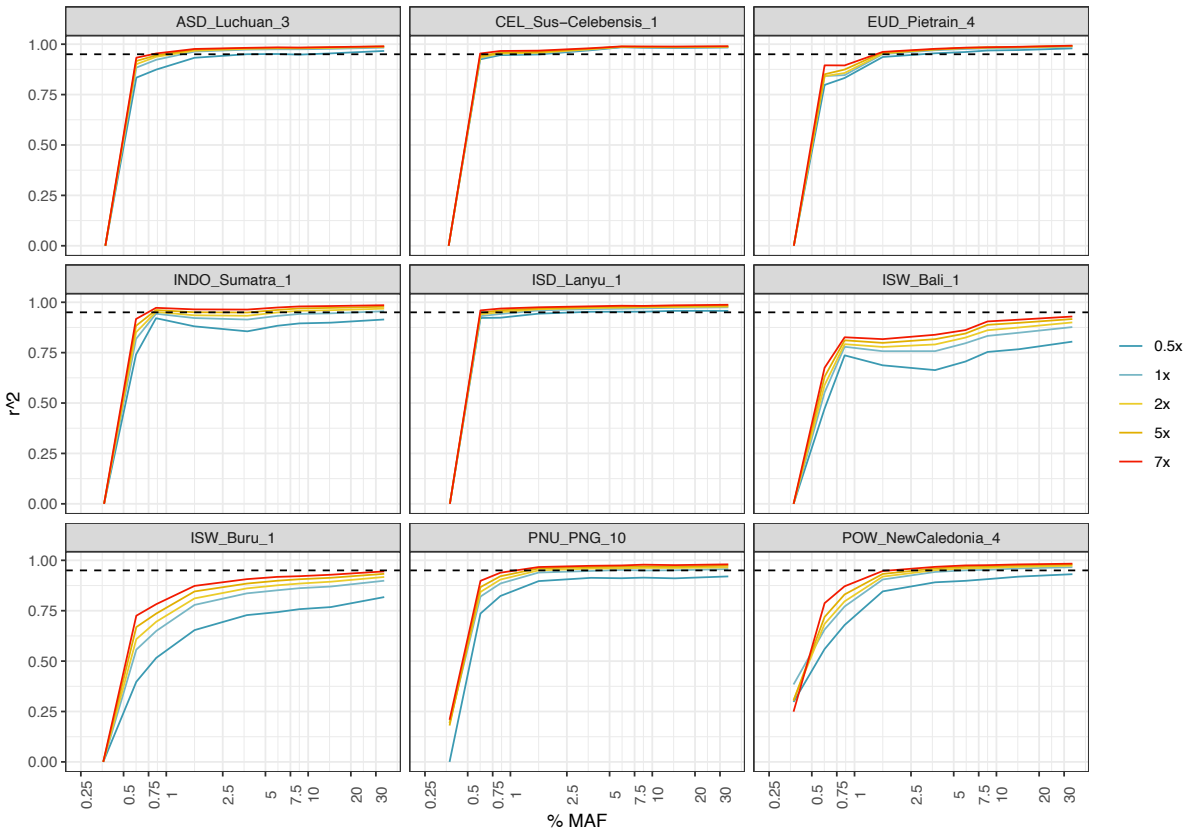


Fig. S4.

Imputation accuracy at different coverage in 9 individuals, ASD_Luchan_4: Southeast Chinese domestic pig (*S. scrofa*), CEL_Sus-Celebensis_1: *S. celebenis*, EUD_Pietrain_4: European domestic pig (*S. scrofa*), INDO_Sumatra_1: Sumatra wild pig (*S. scrofa*), ISD_Lanyu_1: Lanyu (domestic) pig (*S. scrofa*), ISW_Bali_1: Bali wild pig (*S. scrofa*), ISW_Buru1: Buru (Wallacea), wild pig (*S. scrofa*), POW_NewCaledonia_4: New Caledonian wild pig (*S. scrofa*). The x-axis represents MAF (minor allele frequency), y-axis represents concordance between "true" genotypes and imputed genotypes (r^2) at different downsampled coverage (0.5-7x) as computed by the GLIMPSE_concordance tool.

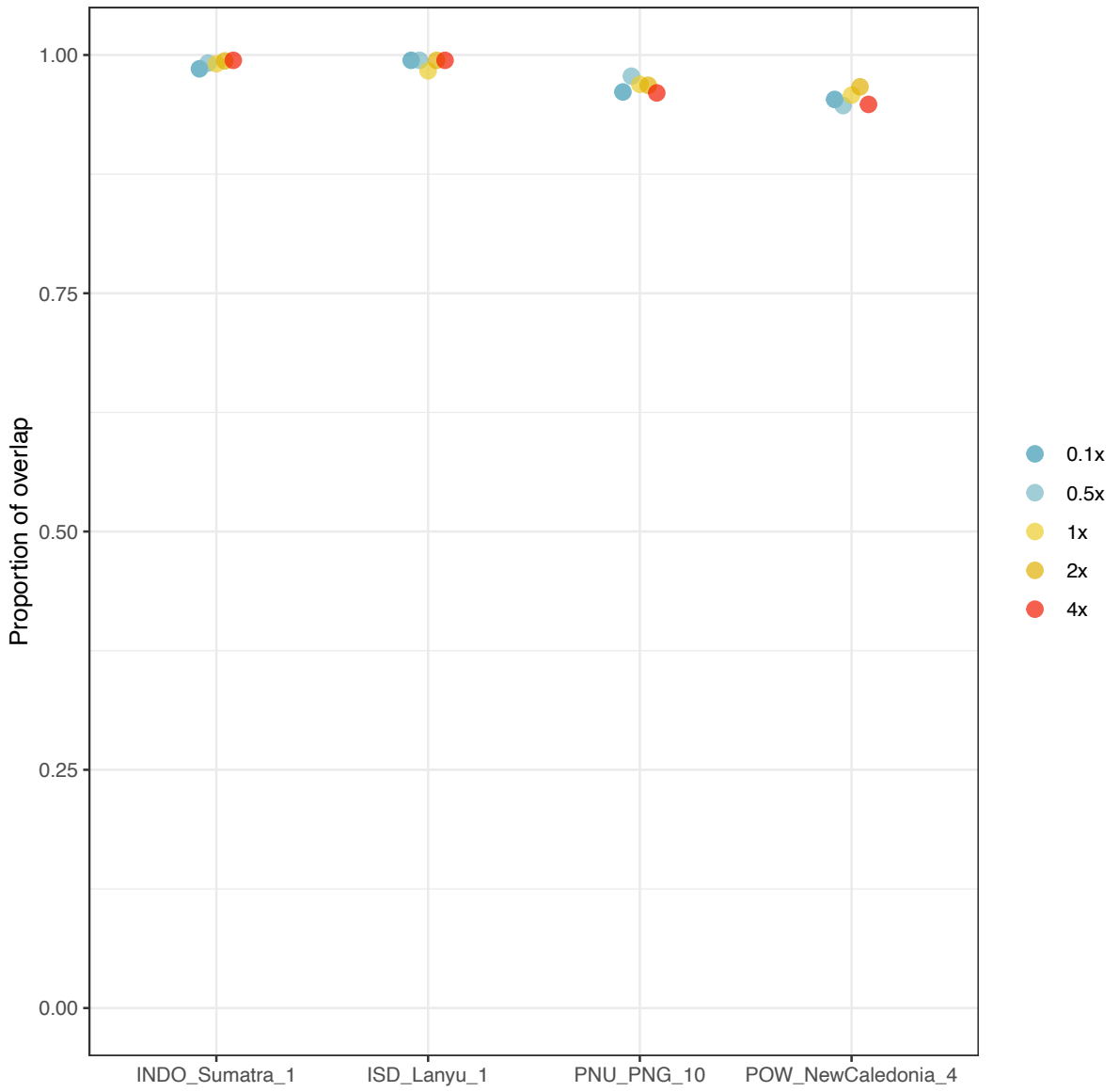


Fig. S5.

Accuracy of local ancestry inference across different sequencing coverages. The y-axis shows the proportion of matching ancestry calls by GNOMIX between high-coverage data and downsampled data 0.5x to 4x coverage.

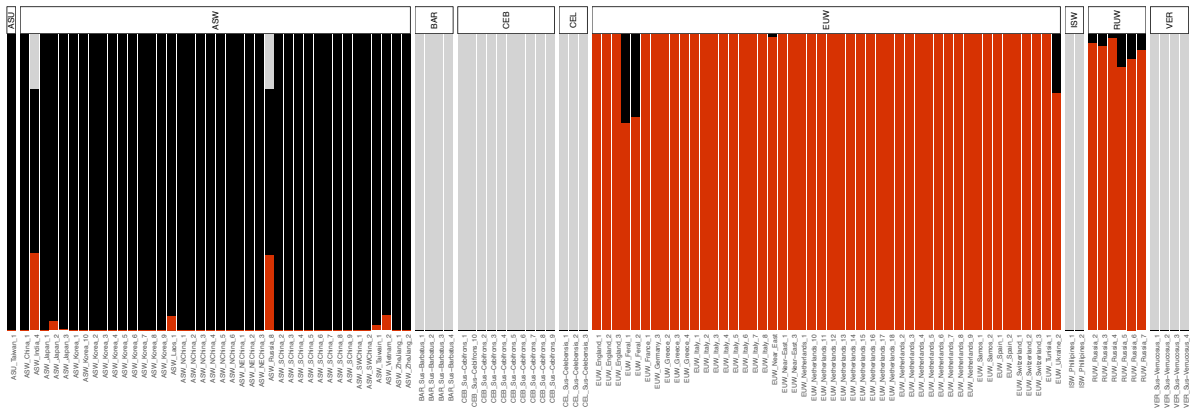


Fig. S6.

Results of unsupervised ADMIXTURE analysis using K=3 based 120 wild boar genomes.

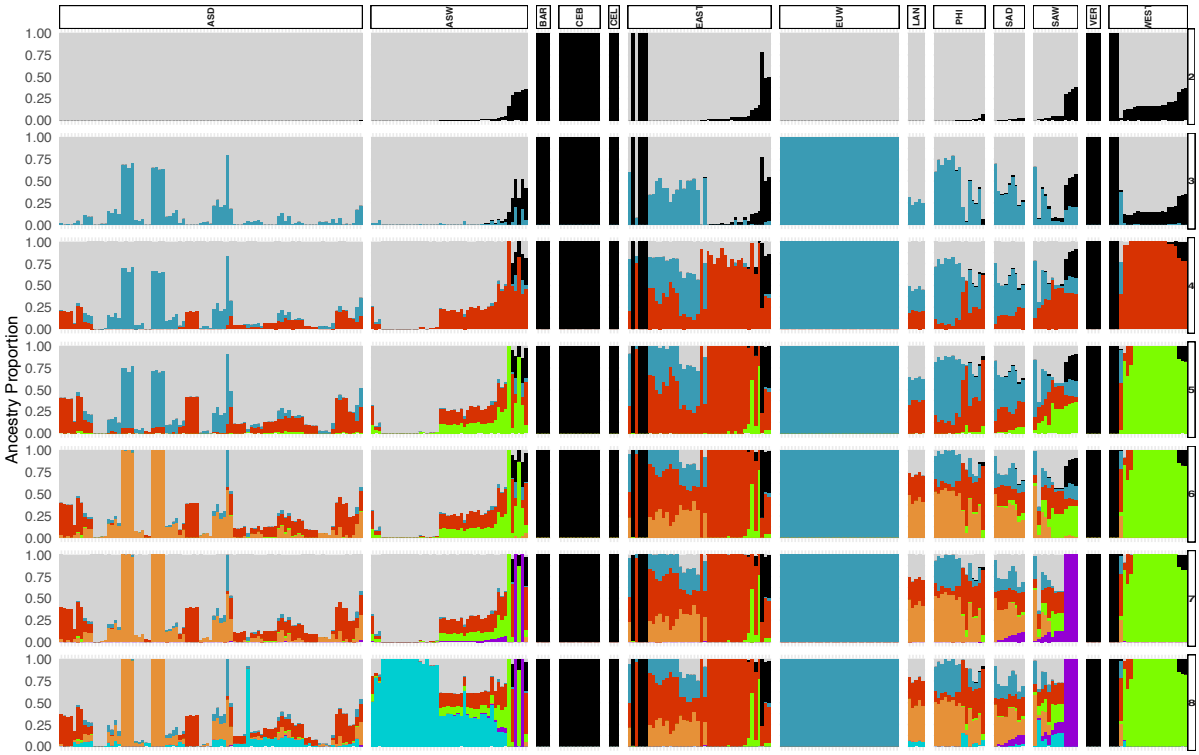


Fig. S7.

Results of unsupervised ADMIXTURE analysis using K=2-8 based on 304 genomes. The following acronym were used to separate individuals into groups: ASD=mainland East Asian domestic, ASW=mainland East Asian , BAR=*S. barbatus*, CEB=*S. cebifrons*, CEL=*S. celebensis*, EAST=ISEA East of the Wallace Line, EUW=European wild, LAN=Lanyu, PHI=domestic pigs from the Philippines, SAD=South Asia domestic (e.g. Bhutan and Nepal), SAW=South Asian wild (e.g. Buthan, India, Nepal), VER=*S. verrucosus*, WEST=ISEA West of the Wallace Line.

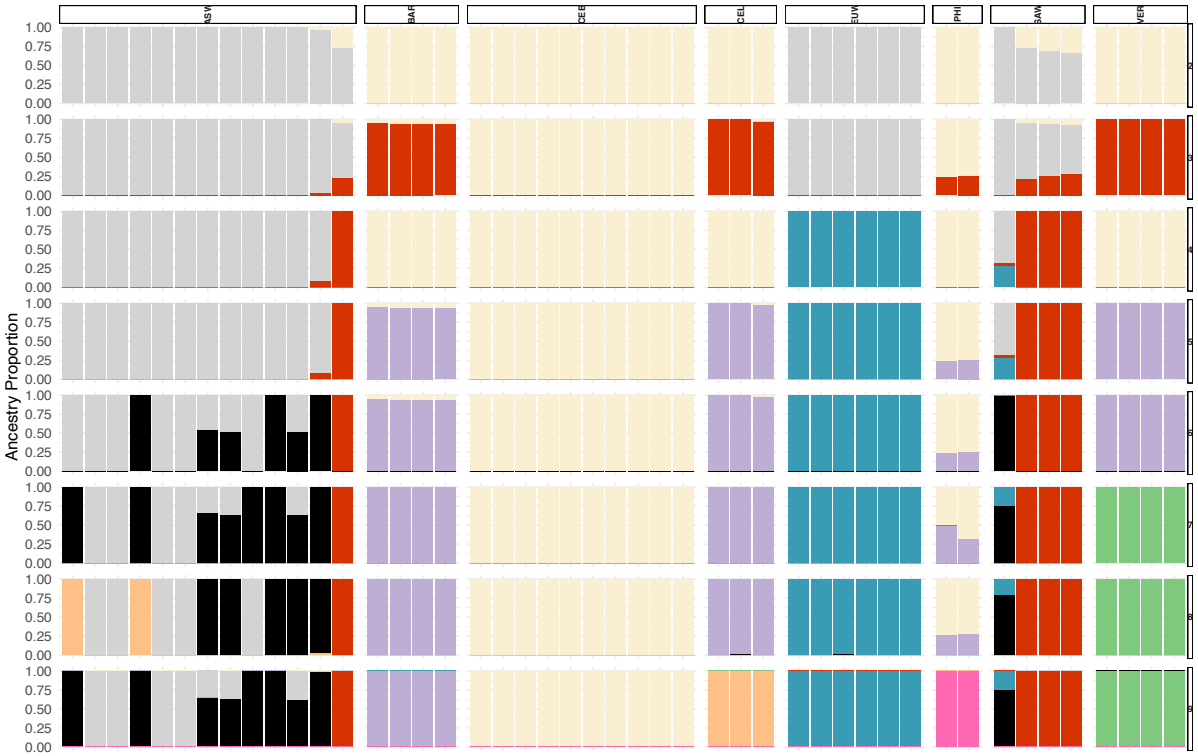


Fig. S8.

Results of unsupervised ADMIXTURE analysis using K=2-8 based on 46 genomes. The following acronym were used to separate individuals into groups: AASW=mainland East Asian , BAR=*S. barbatus*, CEB=*S. cebifrons*, CEL=*S. celebensis*, PHI=S. Philippensis, SAW=South Asian wild (e.g. Buthan, India, Nepal), VER=*S. verrucosus*.

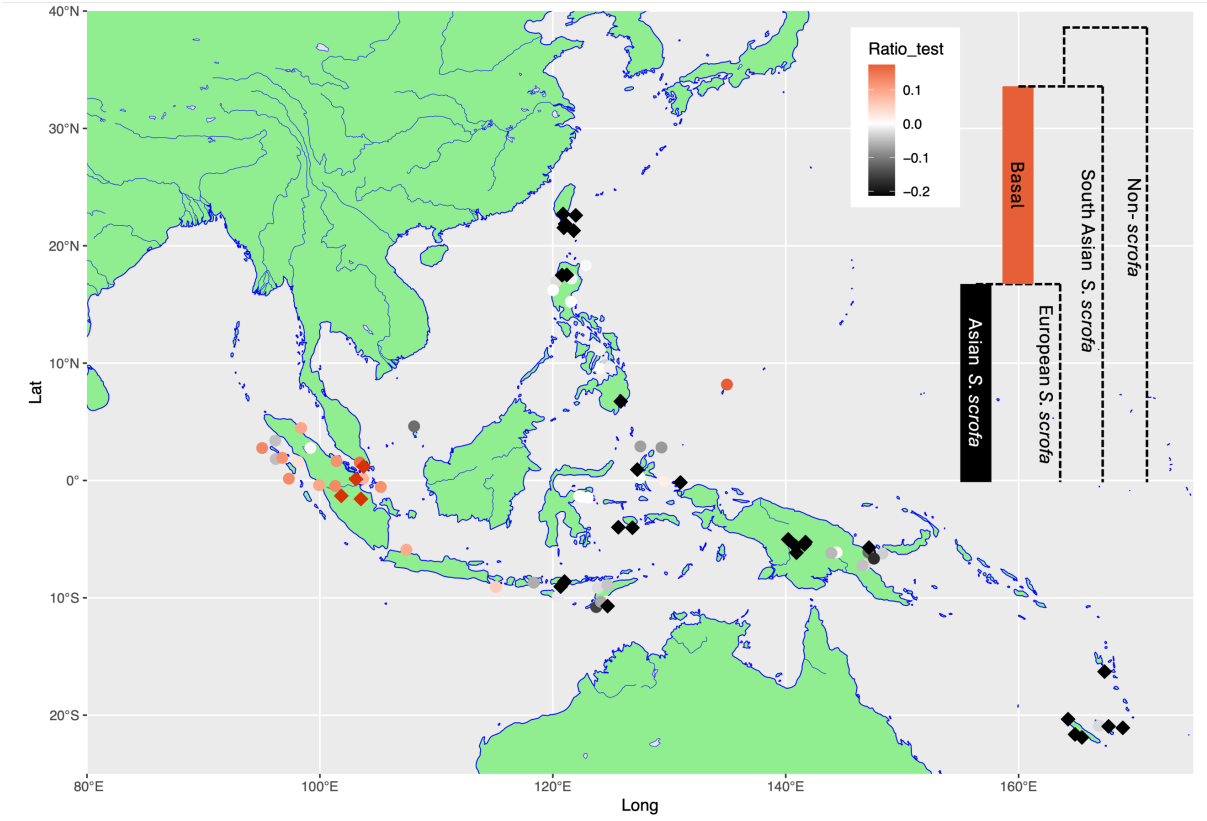


Fig. S9.

Figure showing the fit of admixture graphs assuming basal *S. scrofa* ancestry versus Asian *S. scrofa* ancestry for each sample, accounting for non-*S. scrofa* admixture. Models supporting basal *S. scrofa* ancestry are shown in red, and those supporting Asian ancestry in black. Instances when this support is significant at $p<0.05$ is shown by diamonds.

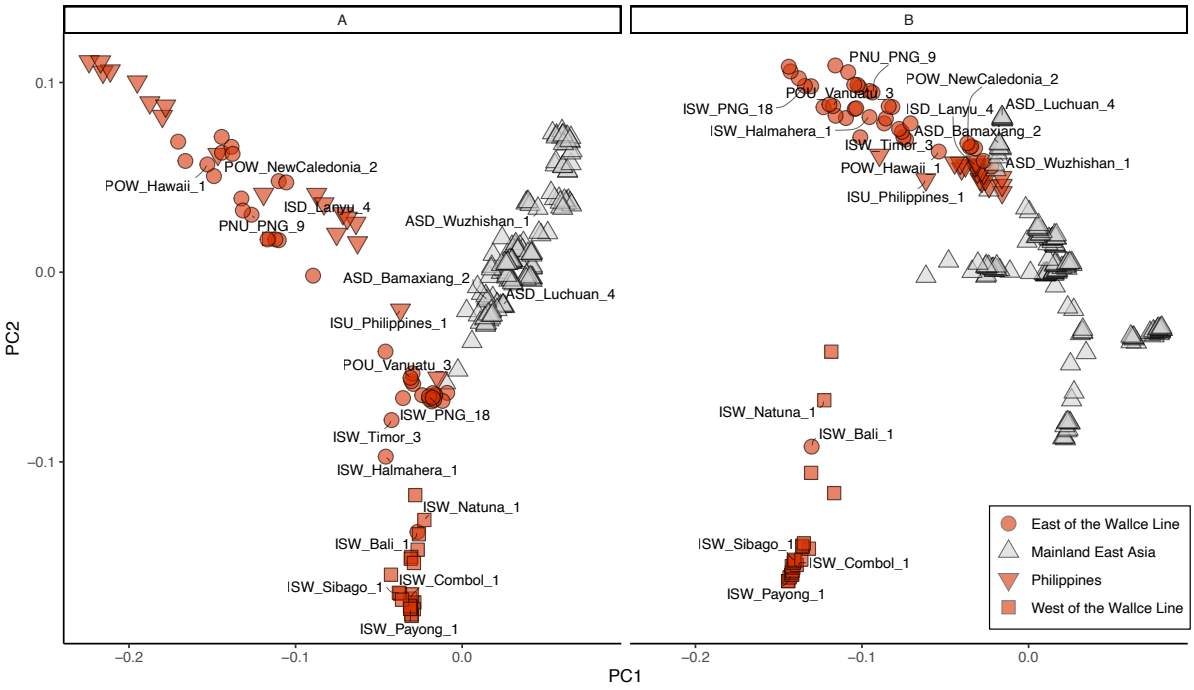


Fig. S10

Principal component analysis using Asian pigs (both domestic and wild) whole genome (**A**) and deconvoluted genomes (**B**), i.e. without ISEA endemic-Sus and European *S. scrofa* ancestry.

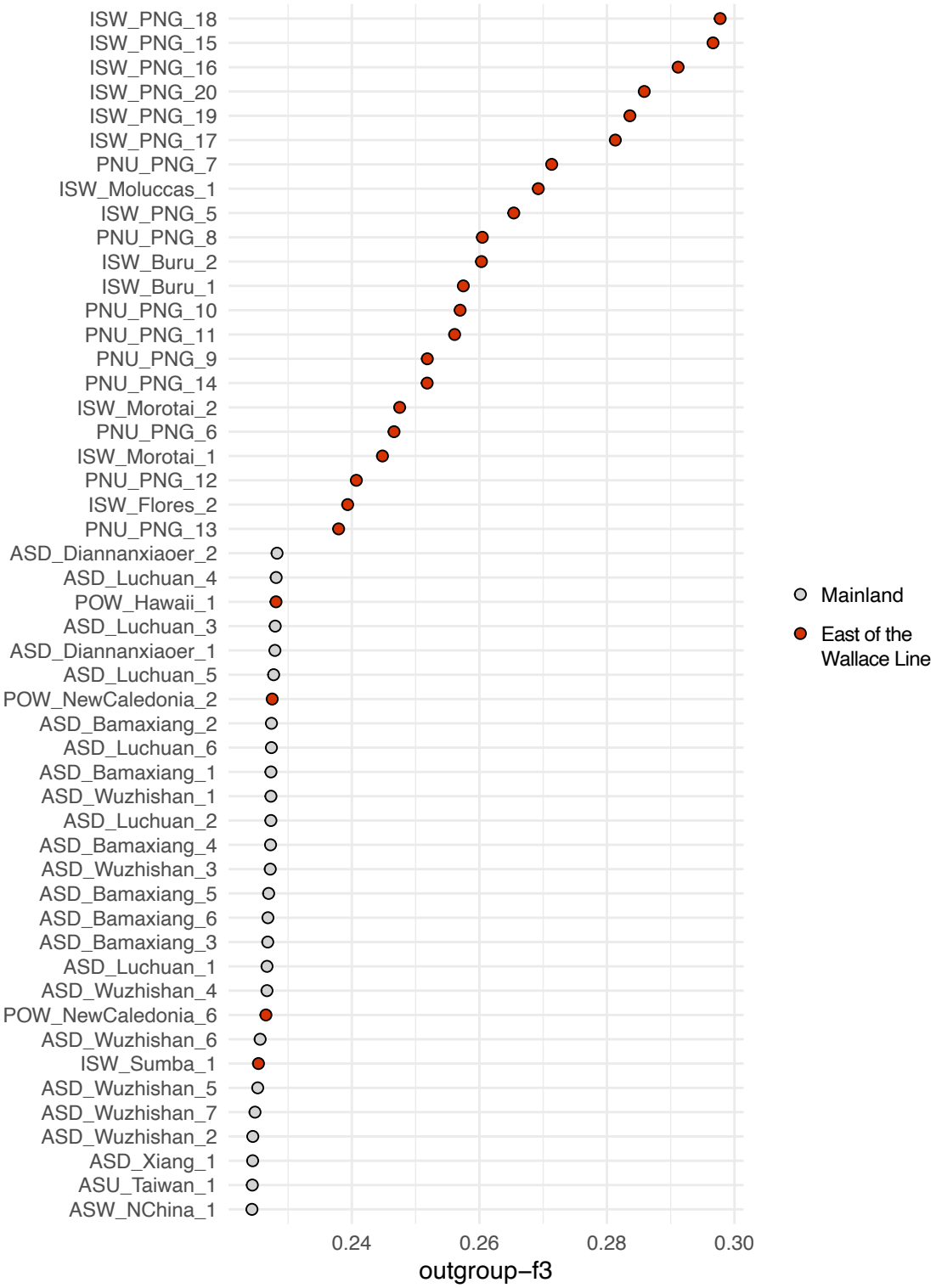


Fig. S11.
The top 50 outgroup-f3 of the form (Pygmy hog [outgroup],X,POU_Vanuatu_3), where X represents any genome in our dataset (y-axis). While this statistic was computed across all individual genomes in our dataset, we have plotted only the top 50 for clarity. POU_Vanuatu_3 represents a 2,500-year-old ancient Vanuatu genome excavated from a Lapita archaeological complex.

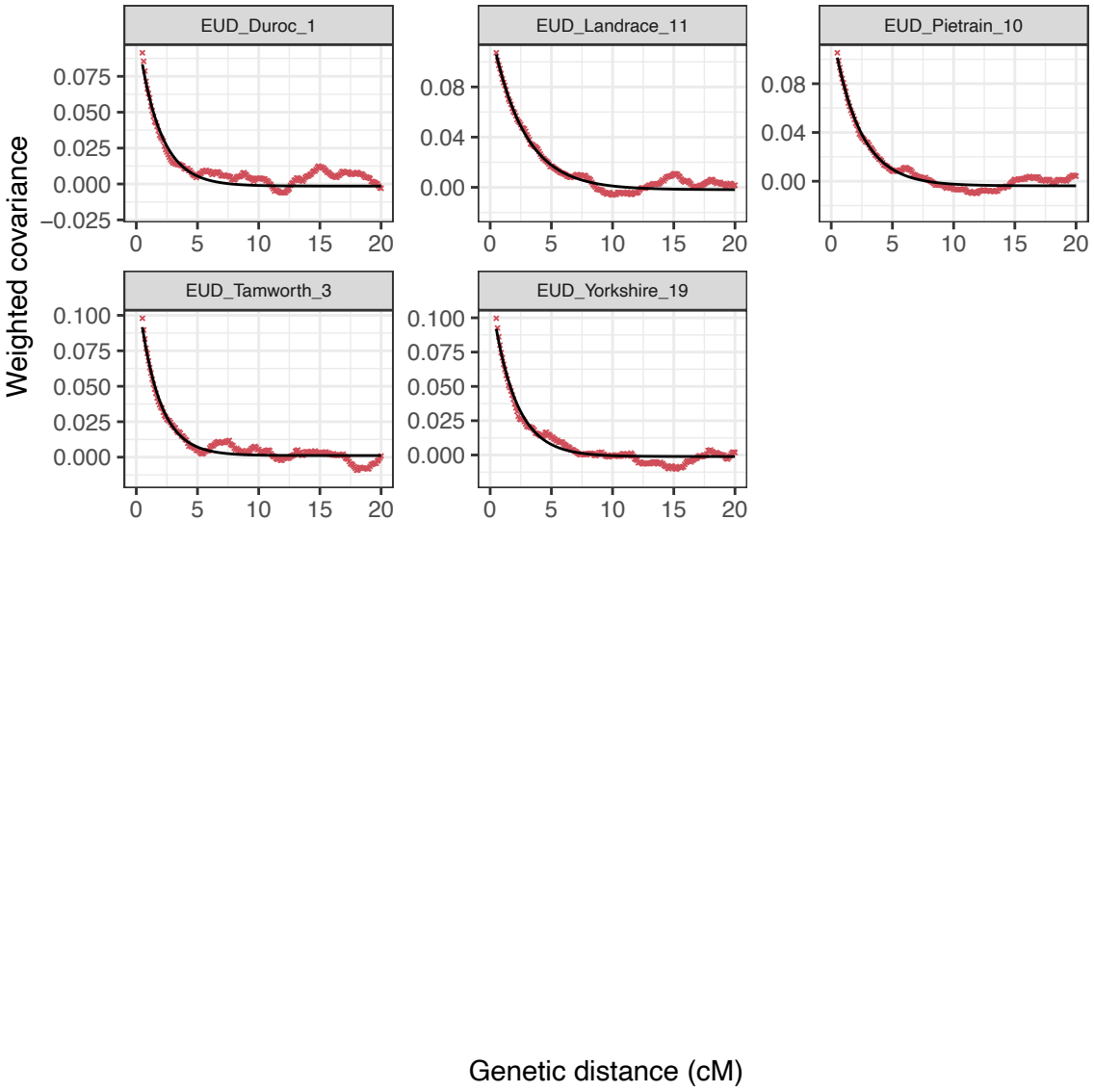


Fig. S12.
Example of exponential function fitted by DATES to ancestry covariance in European pigs which are known to possess Asian ancestry.

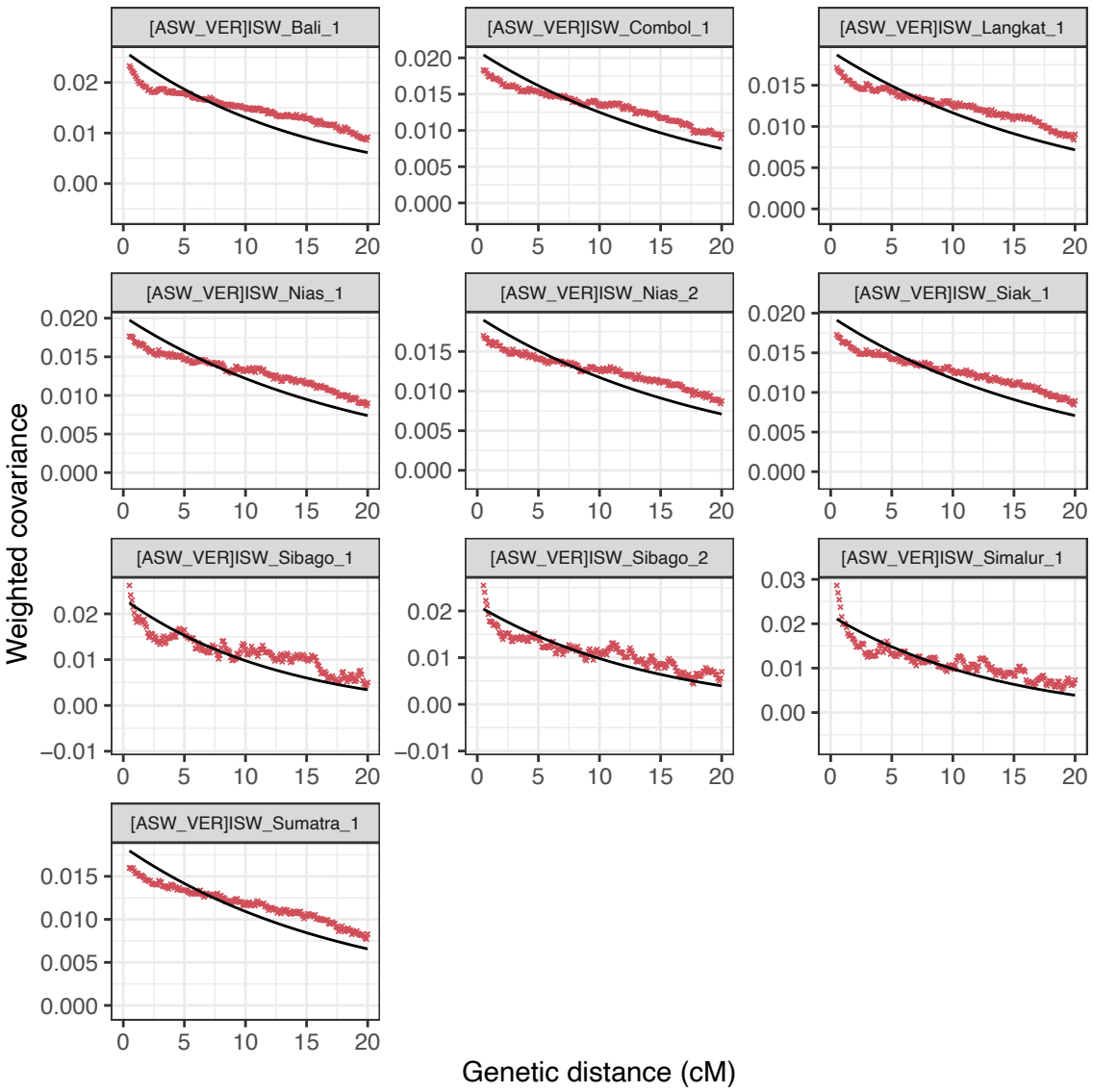


Fig. S13.

Exponential function fitted by DATES to ancestry covariance in Asian ancestry which are known to possess *S. verrucosus* ancestry.

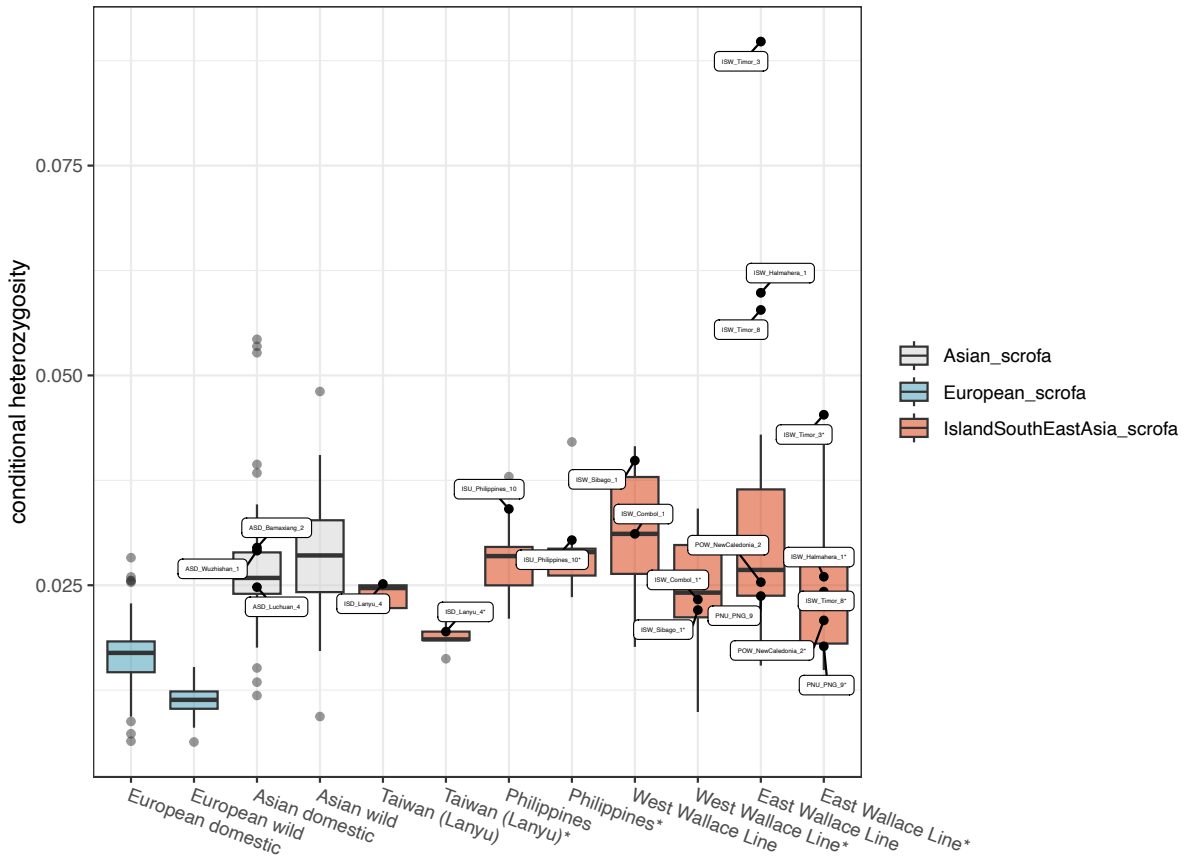


Fig. S14.

(Conditional) heterozygosity estimated by sampling two reads sites found heterozygous in a high coverage *S. cebifrons*. Individuals were grouped based on their location / population. Individuals / populations (on the x-axis) with a * indicate that conditional heterozygosity was computed only at sites that overlap with Asian *S. scrofa* ancestry as per our deconvolution procedure (see e.g. Fig. 3B)

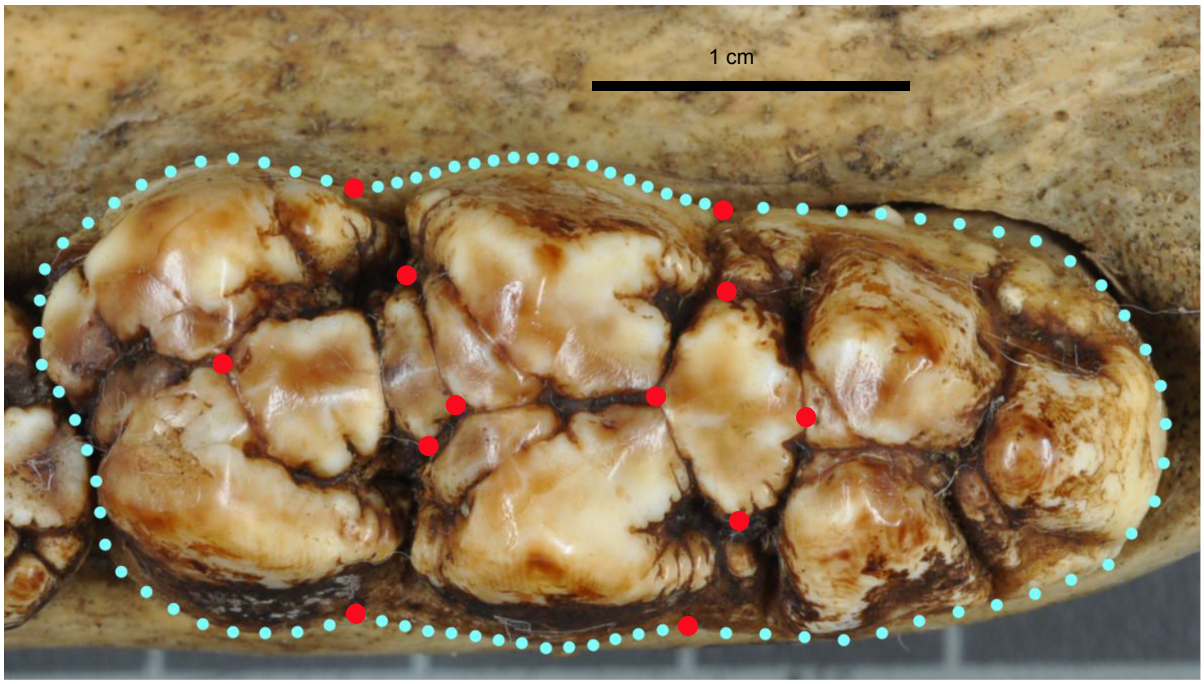


Fig. S15.
Position of the 12 2D-landmarks and 87 sliding semi-landmarks on a third lower molar (specimen nhm-69518).

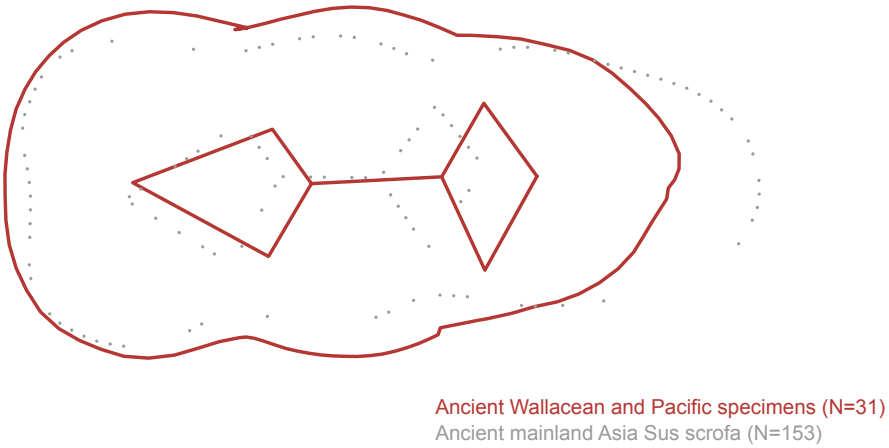


Fig. S16.

Molar shape differences between Wallacean and Pacific specimens (N=31) and those from mainland Asia (N=147). Visualisation of the third lower molar shape variation along the discriminant axis between the two groups.

Data S1. (separate file)
Information about samples newly sequenced data and publicly available data used in this study including provenance, sampling date, age and sequencing statistics.

Data S2. (separate file)
Information about the mitochondrial data used in Fig. 2A.

Data S3. (separate file)
Metadata and contextual information for the specimens morphometrically analyzed in this study. The table includes the name of the institution hosting the specimen, individual identifier (Individual.ID) , taxonomic classification (Genus, species, subspecies), DNA references, and geographical provenance (Country, Site Name, Site Code, Latitude, Longitude). Chronological and cultural attributions are indicated by Chrono, Period, and calibrated dates (Dating BC). Additional contextual data include sex, excavation details (Contact, Feature, Square, Date), inventory numbers (Inventory, Element), notes, the identifications in Fig. 1C and the final morphometric identifications (Supplementary Material).