Ancient DNA reveals genetic admixture in China during tiger evolution

Check for updates

Xin Sun [1,16], Yue-Chen Liu [1,17,18], Mikhail P. Tiunov [2], Dmitry O. Gimranov[3,4], Yan Zhuang[1], Yu Han[1], Carlos A. Driscoll [5], Yuhong Pang[6], Chunmei Li[6], Yan Pan[7], Marcela Sandoval Velasco [8], Shyam Gopalakrishnan[8], Rui-Zheng Yang[1], Bao-Guo Li[9], Kun Jin[10], Xiao Xu[1], Olga Uphyrkina [2], Yanyi Huang [6,11,12], Xiao-Hong Wu[7], M. Thomas P. Gilbert [8,13], Stephen J. O'Brien[14] ✉, Nobuyuki Yamaguchi [15] ✉ & Shu-Jin Luo [1] ✉

The tiger (*Panthera tigris*) is a charismatic megafauna species that originated and diversified in Asia and probably experienced population contraction and expansion during the Pleistocene, resulting in low genetic diversity of modern tigers. However, little is known about patterns of genomic diversity in ancient populations. Here we generated whole-genome sequences from ancient or historical (100–10,000 yr old) specimens collected across mainland Asia, including a 10,600-yr-old Russian Far East specimen (RUSA21, 8× coverage) plus six ancient mitogenomes, 14 South China tigers (0.1–12×) and three Caspian tigers (4–8×). Admixture analysis showed that RUSA21 clustered within modern Northeast Asian phylogroups and partially derived from an extinct Late Pleistocene lineage. While some of the 8,000–10,000-yr-old Russian Far East mitogenomes are basal to all tigers, one 2,000-yr-old specimen resembles present Amur tigers. Phylogenomic analyses suggested that the Caspian tiger probably dispersed from an ancestral Northeast Asian population and experienced gene flow from southern Bengal tigers. Lastly, genome-wide monophyly supported the South China tiger as a distinct subspecies, albeit with mitochondrial paraphyly, hence resolving its longstanding taxonomic controversy. The distribution of mitochondrial haplogroups corroborated by biogeographical modelling suggested that Southwest China was a Late Pleistocene refugium for a relic basal lineage. As suitable habitat returned, admixture between divergent lineages of South China tigers took place in Eastern China, promoting the evolution of other northern subspecies. Altogether, our analysis of ancient genomes sheds light on the evolutionary history of tigers and supports the existence of nine modern subspecies.

The radiation of modern Felidae began with the divergence of an ancient lineage that eventually gave rise to today's five roaring cat species[1], specifically the lion (*Panthera leo*), jaguar (*P. onca*), snow leopard (*P. uncia*), leopard (*P. pardus*) and tiger (*P. tigris*). Unlike some congeneric species that spread out to become the apex predators in other continents, the tiger originated, evolved and remained in the jungle of Asia[1,2]. Fossil records of probable tigers date back to the Early Pleistocene at c.2.0 Ma from northern China and Java (Indonesia)[3–7] and possibly to the

Pliocene to Pleistocene boundary at c.2.6–2.2 Ma if a tiger-like felid, *Panthera zdanskyi*[6], is considered a tiger. Despite its once widespread presence, the glacial and climate oscillations during the Middle to Late Pleistocene probably caused the tiger range to repeatedly contract and expand[5,8–10], resulting in the species' relatively low genomic diversity and a mitochondrial coalescence time of living tigers of a maximum of 110,000 yr ago[11–16]. The recent coalescence time points to a small long-term effective population size for tigers and suggests that ancient tigers might have harboured genetic diversity lost from contemporary tigers.

It is possible that waves of expansions after the Late Pleistocene contraction, followed by isolation across a heterogeneous landscape, have contributed to the present biogeographical pattern of tigers[9,10]. Historically, nine tiger subspecies inhabited a vast region west to the Caspian and Aral Seas, east to Northeast Asia and south to the Sunda Islands[5,8] (Fig. 1; see Extended Data Table 1 for names and codes). During the last century, however, habitat loss, fragmentation and hunting reduced the number of free-ranging tigers from probably over 100,000 to fewer than 4,000 today[17]. The Bali (BAL, *P. t. balica*), Caspian (VIR, *P. t. virgata*) and Javan (SON, *P. t. sondaica*) tigers are all extinct and the South China tiger (AMO, *P. t. amoyensis*) has not been seen in the wild for more than 30 yr[17–19]. Recent genome-wide evolutionary analyses confirmed highly restricted gene flow across the range and supported six living subspecies (Fig. 1), namely, the Amur (ALT, *P. t. altaica*), Indochinese (COR, *P. t. corbetti*), Malayan (JAX, *P. t. jacksoni*), Sumatran (SUM, *P. t. sumatrae*), Bengal (TIG, *P. t. tigris*) and South China tigers[16,20]. Signals of selection have also been revealed in several subspecies, probably associated with local adaptation to various ecosystems[15,20].

Several questions remain if the natural history of the tiger is to be fully resolved. First, the contrast between the species' age and its mitochondrial (mt)DNA coalescence indicates that the genomic diversity and demographic dynamics of ancient tigers might differ from that of their living counterparts. Second, phylogenetic analysis of modern tigers clustered the Sumatran, Javan and Bali tigers into a monophyletic clade, supporting a single ancient colonization event to the Sunda Island[12]; however, the evolutionary pathways of mainland tigers are more complicated. For example, mtDNA gene sequencing suggested that the extinct Caspian tiger is nearly indistinguishable from the Amur tiger[11]. A craniometric analysis showed an extensive overlap between the Caspian tiger and the other mainland subspecies[21], implying an unsettled origin of the Caspian subspecies. Finally, previous population genomic analysis of modern specimens revealed an mtDNA lineage in South China tigers that is basal to all other subspecies and another divergent lineage similar to Indochinese tigers[16,20]. Whether this paraphyly is due to an admixed ancestry in the origin of South China tigers or a poorly defined geographic boundary between Indochinese and South China tigers would require an explicit assessment of voucher specimens of confirmed heritage[9]. Overall, it is likely that genome-scale data from ancient samples will be key to building a complete understanding of the evolution of tigers.

The past decade has witnessed substantial advances in ancient DNA research, now enabling the retrieval of genomic information from extinct mammals that date back to over one million years ago[22–24]. Ancient DNA has been recovered from Felidae including sabre-toothed cats[25–27], the Late Pleistocene Holarctic lions[28–31] and the European Late Pleistocene leopards[32]. However, thus far, only one whole genome[33] has been retrieved from Pleistocene/Early Holocene tiger specimens, partially due to the overlap of the tiger distribution with temperate-tropical forest biomes, where specimens are less likely to be preserved than in colder regions.

To elucidate the more comprehensive evolutionary history of tigers, we collected and generated high-quality genomes from over 60 samples spanning a wide geographic range and timescales, including both zooarchaeological/ancient (that is, 1,000 - 10,000 yr) remains excavated from Northeast Asia and historical (that is, ~100 yr) museum specimens that originated in Central and East Asia (Fig. 1, Extended Data Table 1, and Supplementary Tables 1 and 2). These samples, including one from the Russian Far East (RFE) dated at c.10,600 yr BP and the South China tiger holotype specimen designated in 1905[34], are crucial for understanding tiger dispersal and inhabitation across mainland Asia. Together with published genomes from extant subspecies, we revealed cryptic biogeographic refugia, resolved longstanding controversies in tiger subspecific phylogeny and taxonomy, and for the first time illuminated China as a stepping stone or 'melting pot' for tiger subspecies to establish in mainland Asia.

An optimal conservation strategy for wild tigers may involve establishing corridors or implementing reintroduction programmes to landscapes where tigers have gone extinct. To this end, understanding a species' evolutionary history from an ancient DNA perspective promises to reveal patterns of genomic diversity and connectivity that no longer exist, which in turn may guide today's conservation efforts to preserve the iconic species survived by its living subspecies.

Results and Discussion

Genome dataset of extinct and extant tigers

We retrieved genome sequences from 7 of the 25 ancient RFE felid fossil remains and evaluated their endogenous DNA authenticity (Extended Data Fig. 1 and Supplementary Table 1). The final ancient RFE tiger genome dataset included one whole-genome assembly with 8× coverage from a phalanx bone (RUSA21) excavated from the Letuchaya Mysh Cave and six mitogenomes obtained by RNA probe target capture enrichment (Fig. 1 and Extended Data Table 1). RUSA21 was radiocarbon dated to c.10,600 yr BP and the other six specimens were dated to c.8,600–1,800 yr BP (Supplementary Table 1).
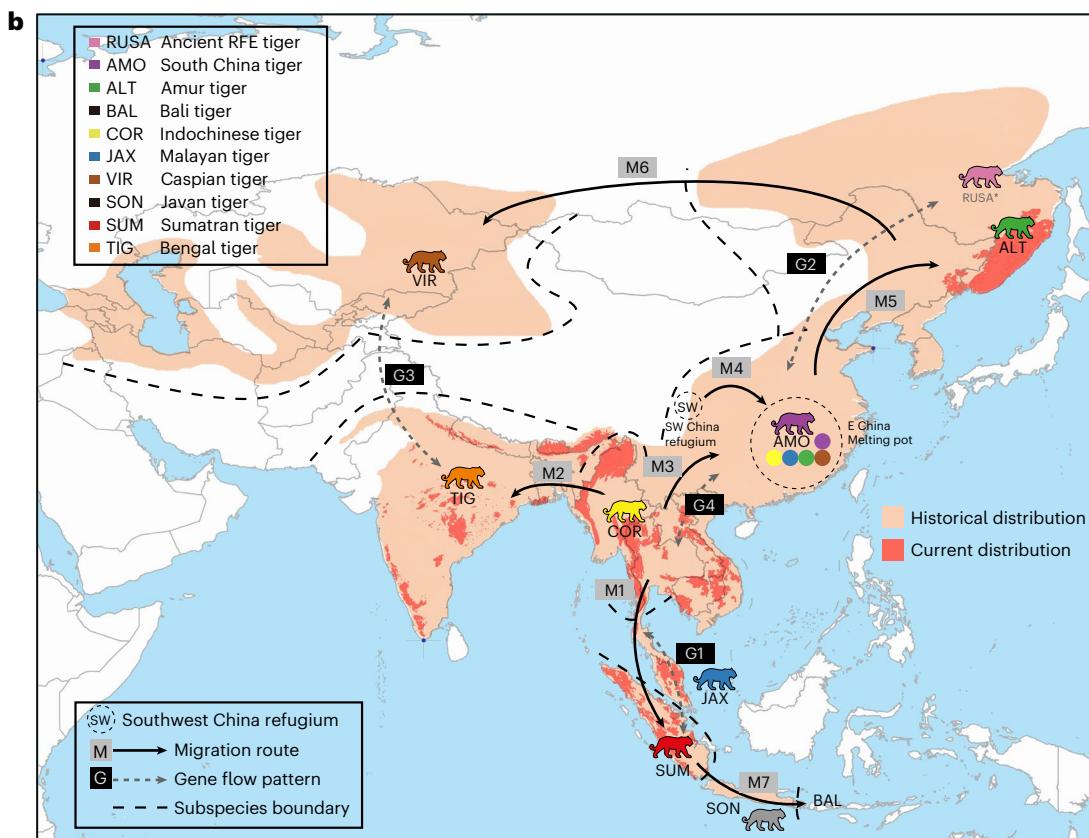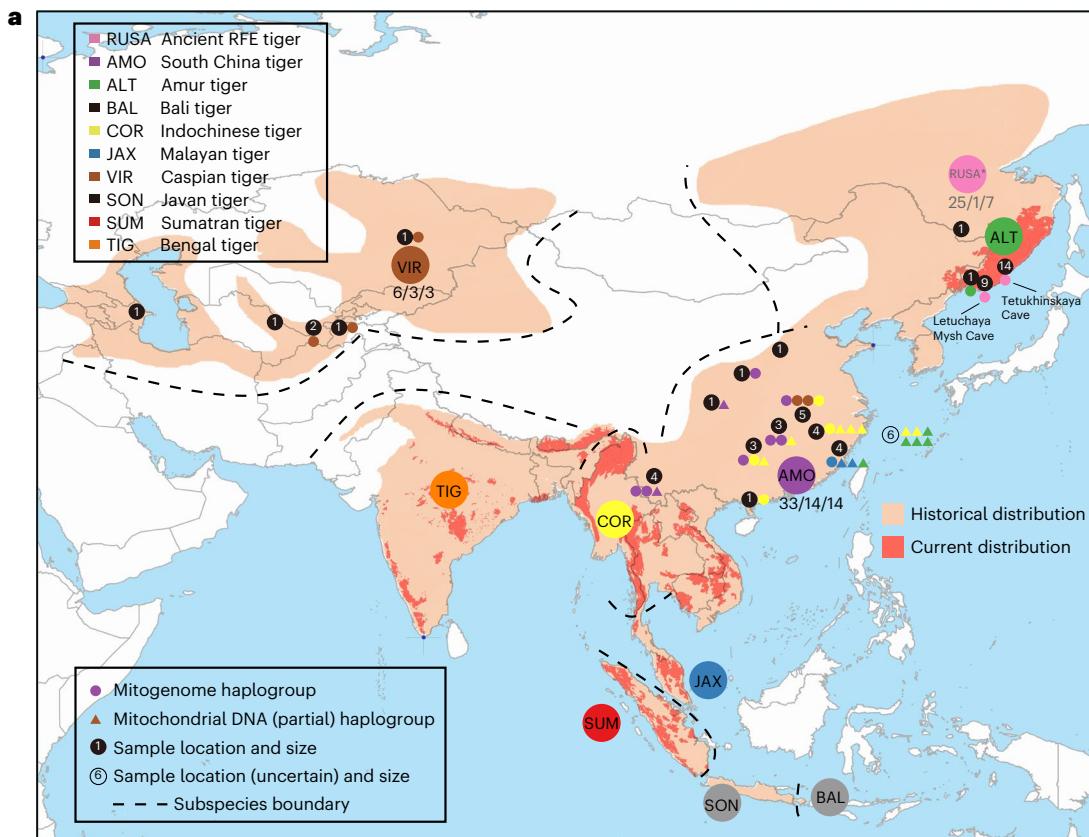
We obtained genome sequencing data over timescales of centuries from tigers recently extirpated from the range, including South China tigers (AMO) and Caspian tigers (VIR), Javan tigers (SON) and Bali tigers (BAL) (Fig. 1, and Supplementary Tables 1 and 2). Eleven South China tiger specimens were sequenced to 1–12×, including one holotype (museum ID 3311) and three paratypes (3305, 3307, 3308) collected in Hankou, Hubei Province, China[34]. Three Caspian tiger specimens originating from Uzbekistan and Kazakhstan were sequenced to 4–8×. We also sequenced 16 South China tiger museum specimens with partial mtDNA fragments encompassing subspecies-diagnostic sites, assembled two mitogenomes using next-generation sequencing data

Fig. 1 | Ancient tiger specimens collected in this study and the inferred dispersal routes of the species based on joint interpretation from modern and ancient tiger genomes. a, Samples and genetic information from the tiger's northern range including China (*N* = 33), the RFE (*N* = 25) and Central Asia (*N* = 6). Small black circles with numbers indicate the geographic origin and sample size from each location. Large coloured circles with abbreviations correspond to the distribution of the nine modern subspecies plus the ancient RFE population. Labels (x/y/z) on the subspecies codes of the South China tiger (*P. t. amoyensis*, AMO), the Caspian tiger (*P. t. virgata*, VIR) and the ancient RFE tigers (RUSA*) correspond to total sample size/the number of whole-genome datasets acquired/the number of mitogenomes assembled. The maternal ancestry of each sample from China and the RFE is colour coded according to its association with modern subspecies or ancient populations and based on either mitochondrial genome (small filled coloured circles) or partial mitochondrial DNA (small filled coloured triangles) sequencing. **b**, The evolutionary routes of tigers reconstructed on the basis of the genome-wide phylogeny, Bayesian molecular dating, modelling of migration and gene flow, and ecological niche modelling of suitable habitats. Seven dispersal routes (M1–M7) and four major post-divergence gene flow events (G1–G4) among tiger populations are indicated. A historical refugium for tigers was postulated in southwest China and the genetic melting pot scenario is illustrated in eastern China, highlighting the significance of mainland China as a stepping stone during tiger expansion.

from Javan and Bali tigers (Fig. 1 and Supplementary Table 2), and incorporated these data into the phylogenetic analysis. The geographic distribution of these samples thus fully covered the historical range of

the extinct tiger subspecies. In combination with 32 previously published whole-genome sequences from living tigers[20], our final dataset consists of 48 nuclear genomes including 15 newly produced genomes
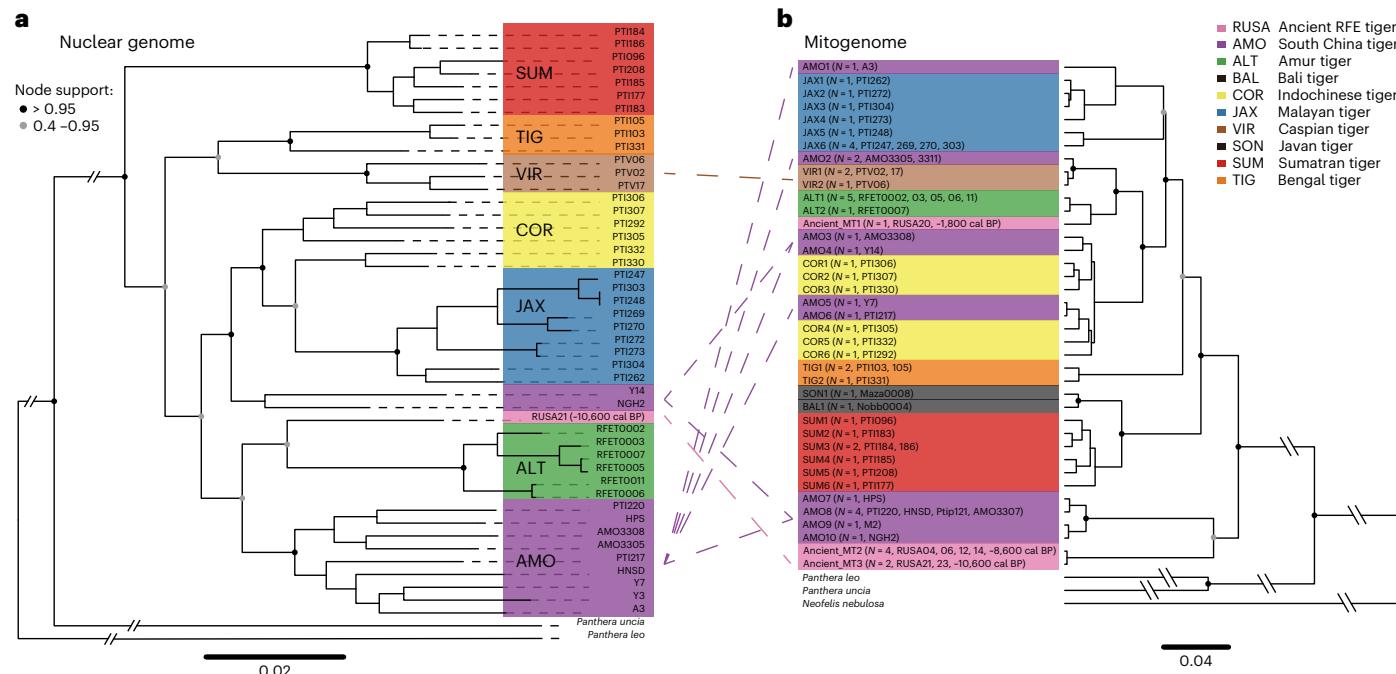
Fig. 2 | Phylogenetic relationships among ancient and modern tigers. Sample labels are colour coded according to their taxonomic classification based on geographical origin. Clades with high bootstrap support are marked on major nodes. **a**, Genome-wide ML phylogeny inferred from 1,242,142 transversions in autosomal neutral regions. Node support was evaluated with 100 bootstrap replicates. Samples are labelled at the tree tips, and the same ancient or historical specimens in the two phylogenetic trees are indicated by dashed lines showing cytonuclear discordance in the South China tiger (*P. t. amoyensis*), the Caspian tiger (*P. t. virgata*) and the ancient RFE tiger (RUSA21). **b**, Bayesian phylogeny inferred from 15,450 BP of mitogenome sequences excluding the control region. Sample labels are mtDNA haplotype codes, followed by the number of samples sharing the haplotype and sample IDs in parentheses. Mitogenome phylogenetic trees based on NJ, MP, ML and Bayesian analyses give identical topologies, and the Bayesian tree is shown here. Nodes with posterior probabilities of 0.4–0.95 and over 0.95 are indicated.

from ancient specimens, and 39 mitogenomes that included 17 novel ones (Supplementary Table 2).

## Distinct subspecies partitioning with cytonuclear discordance

A whole-genome phylogeny (Fig. 2a) was reconstructed on the basis of over 1.2 million transversions (Extended Data Figs. 2 and 3) that were identified from the autosomal neutral regions. In accordance with previous studies, the Sumatran Island subspecies was the earliest branch, followed by the divergence of the Bengal tiger into the Indian subcontinent, the expansion of the Amur tiger from China and Indochina, and the most recent divergence occurring between the Indochinese tiger and Malayan tiger.

In addition, of the five subspecies that were highly distinct[16,20], the newly assembled genomes of the extinct South China tiger and the Caspian tiger each formed a monophyletic clade (Fig. 2b), thus providing robust genome-wide evidence consistent with the subspecies classification and taxonomic designation. The genome-wide autosomal phylogeny placed 9 of the 11 South China tiger (AMO) specimens into a highly supported clade, which then aligned with Amur tigers. The two outlier South China tiger specimens (Y14 and NGH2) clustered as an outgroup to the Southeast Asian clades (COR and JAX, Fig. 2a), probably reflecting their origin from Southwest China in the contact zone adjacent to Indochinese tigers. A group consisting of all Caspian tigers formed a sister clade to the Bengal tiger branch, although bootstrap support for the Caspian–Bengal tiger association was not statistically significant (Fig. 2a).

Intriguingly, mitogenome phylogeny revealed a different evolutionary history, with cytonuclear discrepancies exhibited in all our samples from extinct tigers (Fig. 2b). In contrast to the Caspian–Bengal tiger association indicated by the autosomal phylogeny, the two Caspian

tiger haplotypes clustered with one South China tiger haplotype and then Amur tigers (Fig. 2b). Instead of monophyly, we detected five paraphyletic mitochondrial haplogroups among the 14 South China tiger specimens that either formed a distinct clade basal to all extant tiger subspecies (*N* = 7) or respectively aligned with major mainland tiger haplogroups (that is, Caspian, Malayan and Indochinese tigers), except for the Bengal tiger. In addition, we retrieved partial mtDNA sequences from 16 historical specimens of South China tigers and obtained similar mtDNA diversity based on subspecies-diagnostic sites[20] (Fig. 1 and Supplementary Table 2). In sharp contrast to the autosomal monophyly, mtDNA haplotypes from all 30 South China tigers consistently showed paraphyly, even within the four type specimens. In particular, the holotype specimen (3311), collected from Hankou of Hubei Province in East China, was most closely associated with the Caspian tiger. Since all samples were traceable to geographic origins within the generally recognized South China tiger range, we deduced that the cytonuclear discordance is an intrinsic feature of the natural population consistent with its complex demographic history, rather than artificially introduced admixture in captivity as previously suggested[16,20,35].

Analyses of population genetic structure provided further support for subspecies partitioning in modern tigers (Fig. 3 and Supplementary Table 3). Clear structure of tigers was observed in principal component analysis (PCA; Fig. 3a and Extended Data Fig. 4), including South China tigers and Caspian tigers that each formed their own clusters. ADMIXTURE grouped the tiger genomes according to their subspecies affiliation when the number of groups (*K*) was set to 7 (cross-validation error = 0.79, Fig. 3b). The same two tigers (Y14 and NGH2) from Southwest China that fell within the Indochinese tiger clade in the autosomal phylogeny also showed over half genetic identity related to Indochinese tigers, which is mostly likely attributed to their location in the contact zone. Overall, both Caspian tigers and South China tigers displayed
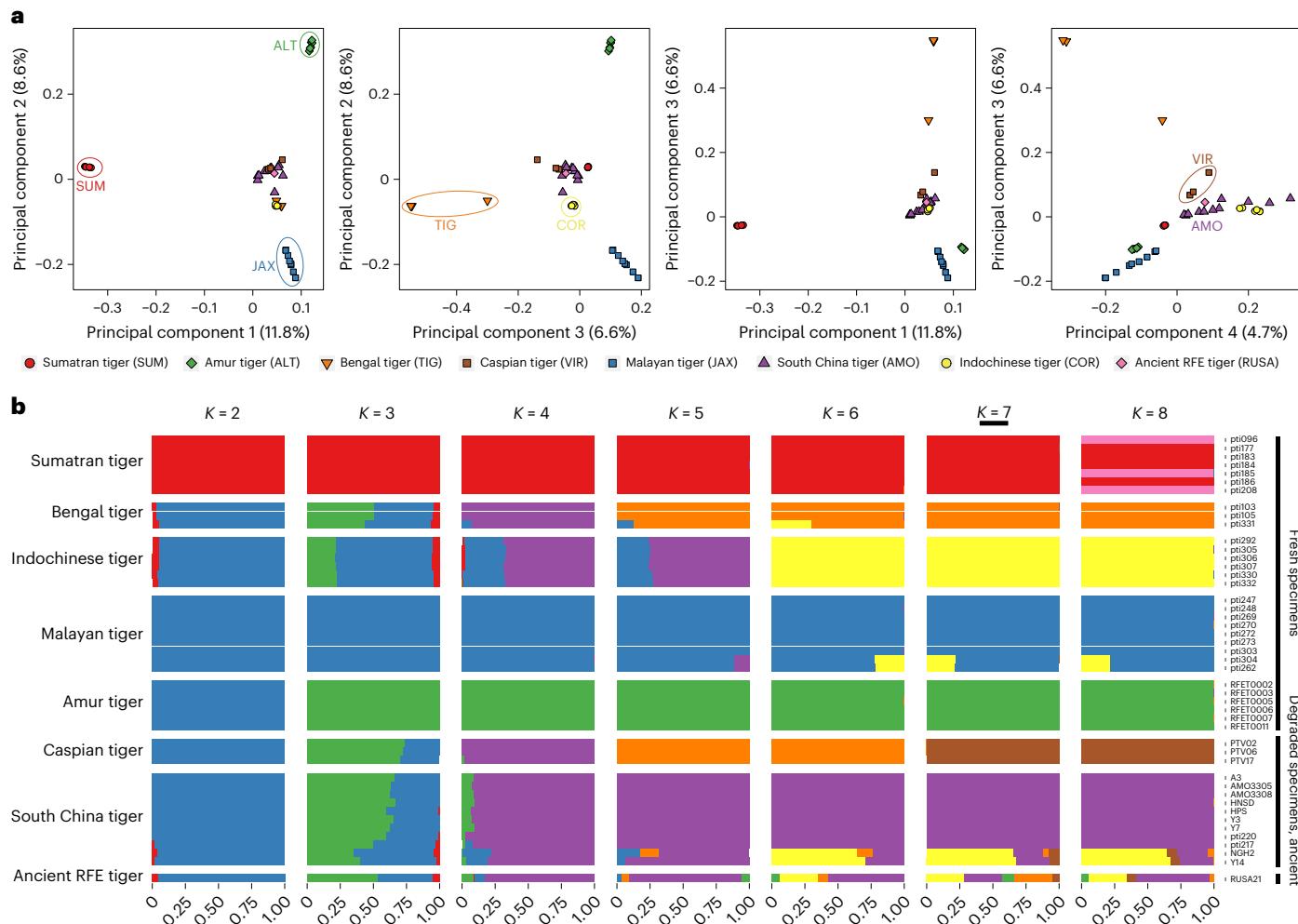
Fig. 3 | Population genomic structure of 13 ancient and 33 modern tigers based on 1,242,142 transversions from autosomal neutral regions. a, PCA of ancient and modern tigers showing the first four principal components. Each tiger is represented by one icon whose colour and shape correspond to the subspecific affiliation of the specimen. **b**, Population genetic structure estimated in ADMIXTURE with the number of inferred populations (*K*) from 2 to 8. Each horizontal bar corresponds to one tiger individual and is proportionally partitioned into *K* coloured segments representing the individual affiliations with each cluster. When *K* = 7, the 7 inferred ancestral populations are in accordance with the 7 highly distinctive modern tiger subspecies. Two *P. t. amoyensis* and the ancient RFE tiger show admixed genomic composition.

nuclear genome-wide phylogenetic and population genomic distinctiveness in support of their traditional classification as subspecies.

Population dynamics of ancient RFE tigers over 10,000 yr
The 10,600-yr-old tiger specimen (sample ID: RUSA21) excavated from the Russian Far East, from which we produced a high-quality ancient genome with 8.1× coverage, possessed genetic affinity not only with modern Amur tigers in the same region but also with other tiger populations further south. Autosomal phylogeny clustered RUSA21 with present-day Amur tigers from the same region, and together they formed the sister clade to the monophyletic South China tigers group (Fig. 2a). In the PCA, however, RUSA21 was associated with South China tigers, not with modern Amur tigers from the same region (Fig. 3a). ADMIXTURE analysis also demonstrated a highly admixed background in RUSA21, including components from nearly all tiger subspecies in mainland Asia, albeit with the largest proportion being South China tiger-like (Fig. 3b).

D-statistics further suggested that the extent of gene flow between a modern subspecies and RUSA21 was minor in comparison with that between two modern subspecies (Fig. 4 and Extended Data Fig. 5), possibly indicative of an affinity of RUSA21 to some outgroup-like, deeply

diverged 'ghost' lineage. Following a nuclear phylogeny framework with ancient (RUSA) and modern Amur tigers (ALT) clustered into an ingroup to all other mainland tigers (PopX), the *D*-statistics inferred an excessive level of allele sharing between Amur tigers and other mainland subspecies relative to that between modern and ancient populations (*D* < 0, Fig. 4b). Alternatively, the significant *D*-statistic value below zero may be interpreted as an excessive amount of allele sharing between the ancient tigers and the outgroup, which may reflect a genetic contribution of an ancestral 'ghost lineage' in East Asia that diverged earlier than modern tigers to the ancient Russian Far Eastern tigers. Although being a singleton sample, the diversity pattern revealed in RUSA21 illuminated a genomic ancestry in the ancient tiger population in the Russian Far East region that is distinct from its modern counterpart.

Our sampled Russian Far Eastern tiger mitochondrial genome sequences spanned different geological periods at c.10,600 yr BP, c.8,600 yr BP, c.1,800 yr BP and the present, illustrating the temporal dynamics of population genomic diversity (Figs. 2b and 5a). The consistency of sequence identity, excavation location and radiocarbon dating data across different samples supported that the seven ancient RFE tiger specimens (RUSA) might belong to three individuals or maternal lineages and correspond to two clusters (Fig. 1, and
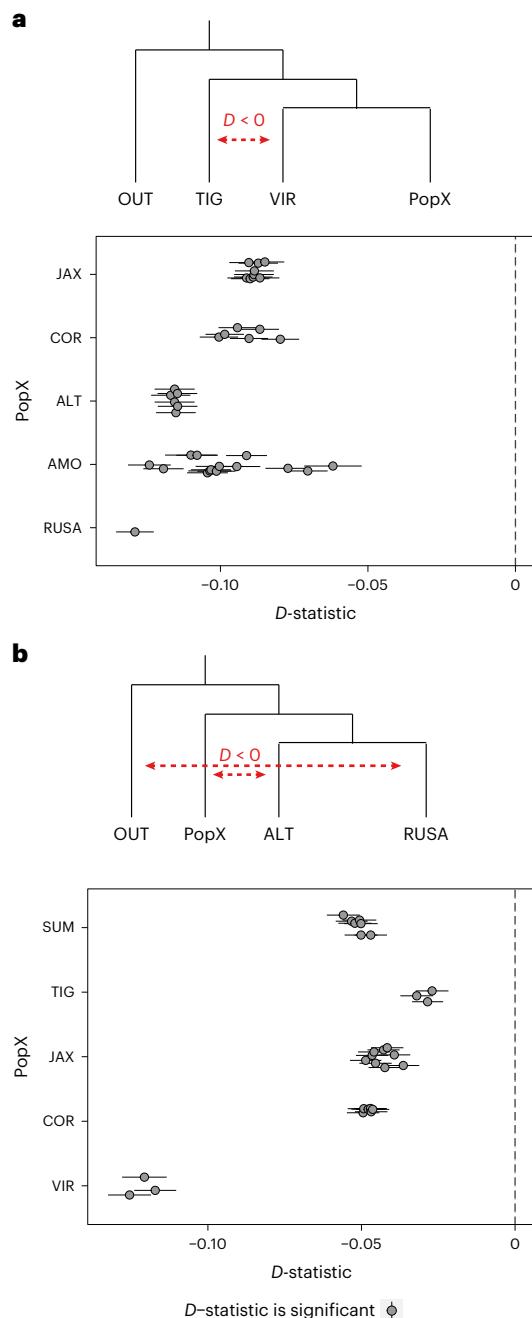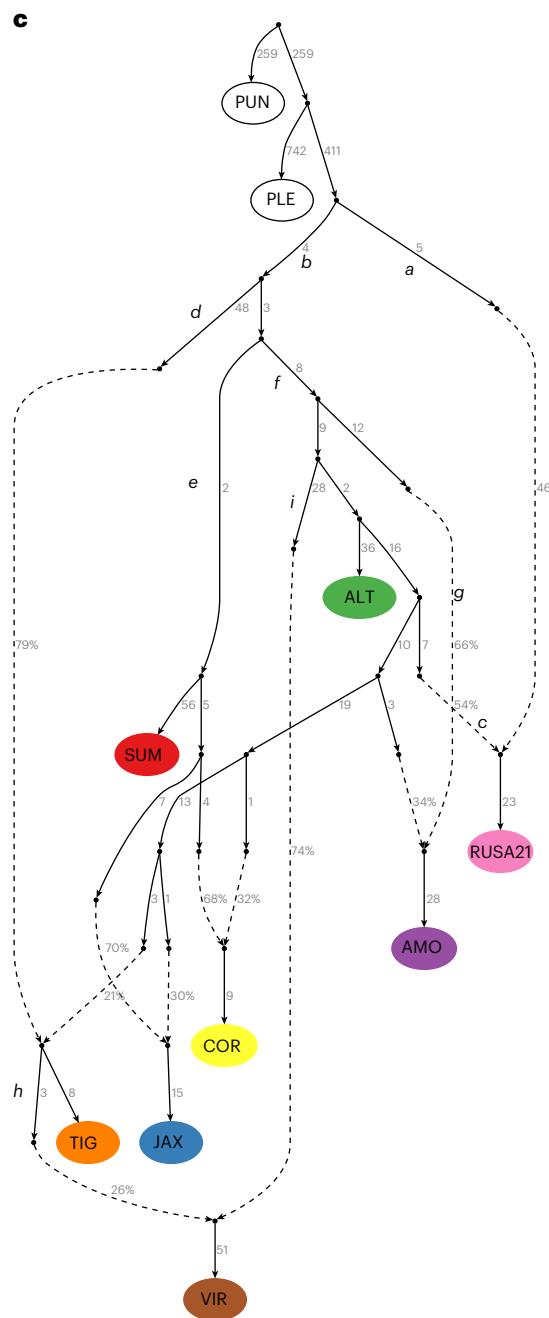
Fig. 4 | The D-statistic inferences for possible scenarios of gene flow and f-statistic-based admixture graph modelling based on ancient and modern specimens. a, The D-statistic testing for excessive allele sharing between the Bengal tiger and the Caspian tiger relative to that between the Bengal tiger and other tiger subspecies from mainland Asia (PopX). **b**, Bottom: the D-statistic testing for excessive allele sharing between the ancient Russian Far Eastern tiger (RUSA) and other mainland tiger subspecies (PopX) relative to that between the Amur tiger and other mainland tiger subspecies (PopX). The D-statistic values are shown on the x axis and tiger samples representing different PopX are shown on the y axis and grouped by subspecies. The D-statistic test is considered significant when the absolute value of transformed z-score |z| > 3. Top: the possible scenario of excessive allele sharing inferred from the D-statistics. Error bars show 1 × s.e.m. from a block jackknife resampling method with a block size of 5 centimorgans. **c**, Admixture graph modelling of the ancestry of tigers with the best-fitting model shown. Solid lines represent phylogenetic relationships among tiger populations and numbers mark branch length in units of genetic drift. Dashed lines show the inferred scenario of genetic admixture events between lineages, with the percentages corresponding to admixture proportions from two ancestral populations. Admixture graphs were built using transversion sites only and one sample with the highest sequencing depth was used to represent each subspecies. PLE, the lion *P. leo*; PUN, the snow leopard *P. uncia*; the rest of the abbreviations for the specimens are as in Fig. 1.

Supplementary Tables 1 and 2). Two closely related mtDNA haplotypes, Ancient_MT2 (shared by four specimens: RUSA04/06/12/14, Tetukhinskaya Cave, c.8,600 yr BP) and Ancient_MT3 (carried by two specimens: RUSA21/23, Letuchaya Mysh Cave, c.10,600 yr BP), distantly coalesced with the unique South China tiger lineage 91,900 yr ago (95% CI: 53,900–130,700 yr ago, Fig. 5a) to form the mtDNA clades basal to all modern tigers. In contrast, a more recent specimen (RUSA20, Letuchaya Mysh Cave, c.1,800 yr BP) carried an mtDNA haplotype (Ancient_MT1) closely associated with extant Amur tigers from the same region, and the two mtDNA lineages shared a time to the most recent common ancestor (TMRCA) at about 16,100 yr ago (95% CI: 6,700–26,200 yr ago, Fig. 5a). Hence, the three ancient and one modern

RFE tiger mtDNA lineages together spanned a chronology of 10,000 yr and may recapitulate part of the population genetic dynamics of tigers on the northeastern edge of the range.

Despite the limited sampling, the temporal transition of mtDNA lineages between c.10,600, c.8,600 yr BP and c.1,800 yr BP is apparent and consistent with a scenario of post-LGM (Last Glacial Maximum) migratory waves from eastern and northern China to the Russian Far East[33,36]. The 10,000-yr-old ancient RFE specimen probably represented one of the earlier waves of tiger expansion to the northeastern edge of the range, based on post-LGM ecological niche modelling (ENM; Extended Data Fig. 6). The genetic ancestry of RUSA21 involved an early-diverging lineage basal to and absent from modern tigers as well as contribution from lineages related to extant Amur tigers (Fig. 4c). ENM-based habitat modelling projected the presence of refugium in Far Eastern Asia (today's Japan Island) during a glacial maximum period that may have sustained relic tiger lineages (Fig. 6 and Extended Data Fig. 6). The probable presence of a refugium in the south of the RFE is also evidenced in the Late Pleistocene cave deposits of new endemic species of pikas and giant flying squirrels[37,38], as well as species that went extinct in the rest of the territory as early as the Early Pleistocene[39,40]. More recently, the evolutionary association between RFE tigers from c.1,800 yr BP to the present provided evidence for the establishment of modern Amur tigers and population genetic continuity in the region over the past two millennia.

The evolutionary origin of Caspian tigers

We elucidated for the first time the evolutionary history of the tiger in its westmost range from a whole-genome perspective, featuring the establishment of the Caspian tiger. The extinct Caspian tiger once occupied a vast area from the riverine and forest habitat in Northwest China to the Caspian Sea with Tugay vegetation[11,41]. In a previous study, based on 20 different museum Caspian tiger specimens originating from Northwest China, Kazakhstan, Tajikistan, Uzbekistan, Turkmenistan, Azerbaijan and Iran, only one major mtDNA haplotype was identified, which differed by only one nucleotide from Amur tigers across a 1.2 kb mitochondrial region; hence, synonymy between these two subspecies was proposed[11]. However, in this study, analysis of the same samples used in ref. 11 at a mitogenome-wide scale revealed close relatedness but distinctiveness of the Caspian and Amur tiger mitogenomes (Fig. 2b). The Caspian tiger mitogenomes (haplotypes VIR1 and VIR2) are most closely associated with two South China tiger type specimens' mitogenomes (haplotype AMO2 shared by specimens 3305 and 3311), which then cluster with Amur tigers haplogroup (haplotypes ALT1 and ALT2) to form a northern mainland Asian clade. In addition, nuclear genome monophyly and the phylogenetic association with the Bengal tiger (Fig. 2a) further suggested a unique evolutionary origin of the Caspian tiger.

Three scenarios concerning the ancient colonization of Central Asia by tigers were previously postulated[11] (Extended Data Fig. 7): (1) a southern route via the Indian subcontinent, (2) a northern route via the Siberian plain and (3) a historical 'Silk Road' route through the Gansu corridor in Northwest China. Phylogenomic analyses and biogeographic modelling supported a possible initial expansion from East Asia to the modern range in Central Asia via the northern Siberian route (Route 2), followed by subsequent gene flow from the ancient Bengal tiger counterpart through the Himalayan corridor (Route 1), a previously proposed biogeographical pass[10].

Although the Caspian tiger shared a recent common ancestor with one South China tiger mitogenome, the lack of a continuous distribution of suitable habitat along the 'Silk Road' from the LGM to the present excluded it (Route 3) as a potential tiger dispersal pathway. The inclusion of distinctive lineages from the Amur, Caspian and South China tigers in one mitochondrial clade (ALT/VIR/AMO, Fig. 2b) supported a once-common ancestor of these mainland subspecies in the northern tiger range, whose ancient radiation centre may have been located in East Asia. One ancient lineage expanded westbound via Siberia in the north, eventually giving rise to the Caspian tiger in Central Asia (Fig. 1 and Extended Data Fig. 7).

To disentangle whether post-divergence gene flow may have contributed to the cytonuclear discordance in the Caspian tiger (mtDNA association with South China tigers vs nuclear DNA affinity with Bengal tigers), D-statistics were applied to evaluate the excessive amount of allele sharing between populations (Fig. 4 and Extended Data Fig. 5). A statistically significant level of excessive allele sharing was detected between Caspian and Bengal tigers (D < 0, Fig. 4a), supporting the hypothesis that the nuclear DNA phylogenetic placement of the Caspian tiger as the sister clade to the Bengal tiger (Fig. 2a) was due to post-divergence gene flows between the two. Such genetic admixture was probably mediated via long-distance, male-biased dispersal that is consistent with the tiger's natural history[42] and the absence of Bengal tiger mtDNA haplotypes in Caspian tigers. These findings offer a direct scientific basis for the reintroduction of tigers into the historic range of the extinct Caspian tigers, and Amur tigers and Bengal tigers may serve as the most related living subspecies and hence possible candidates to regenerate the tiger's evolutionary history.
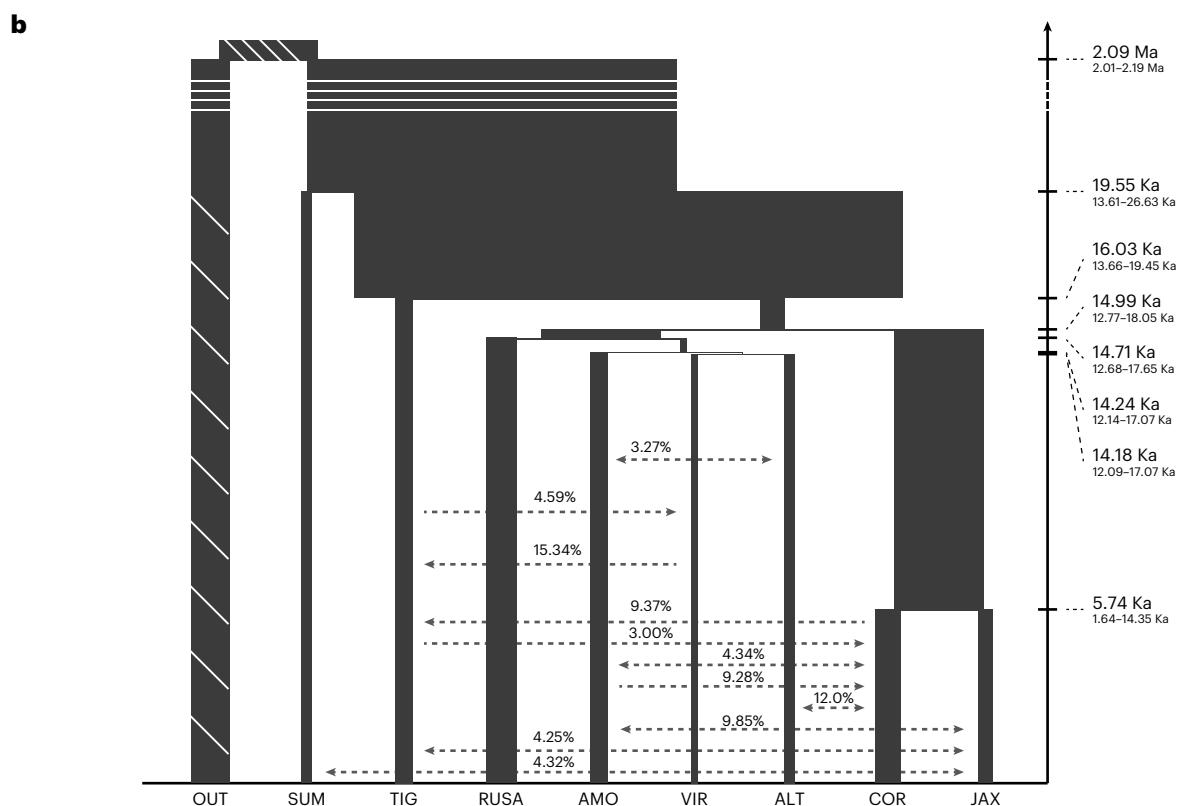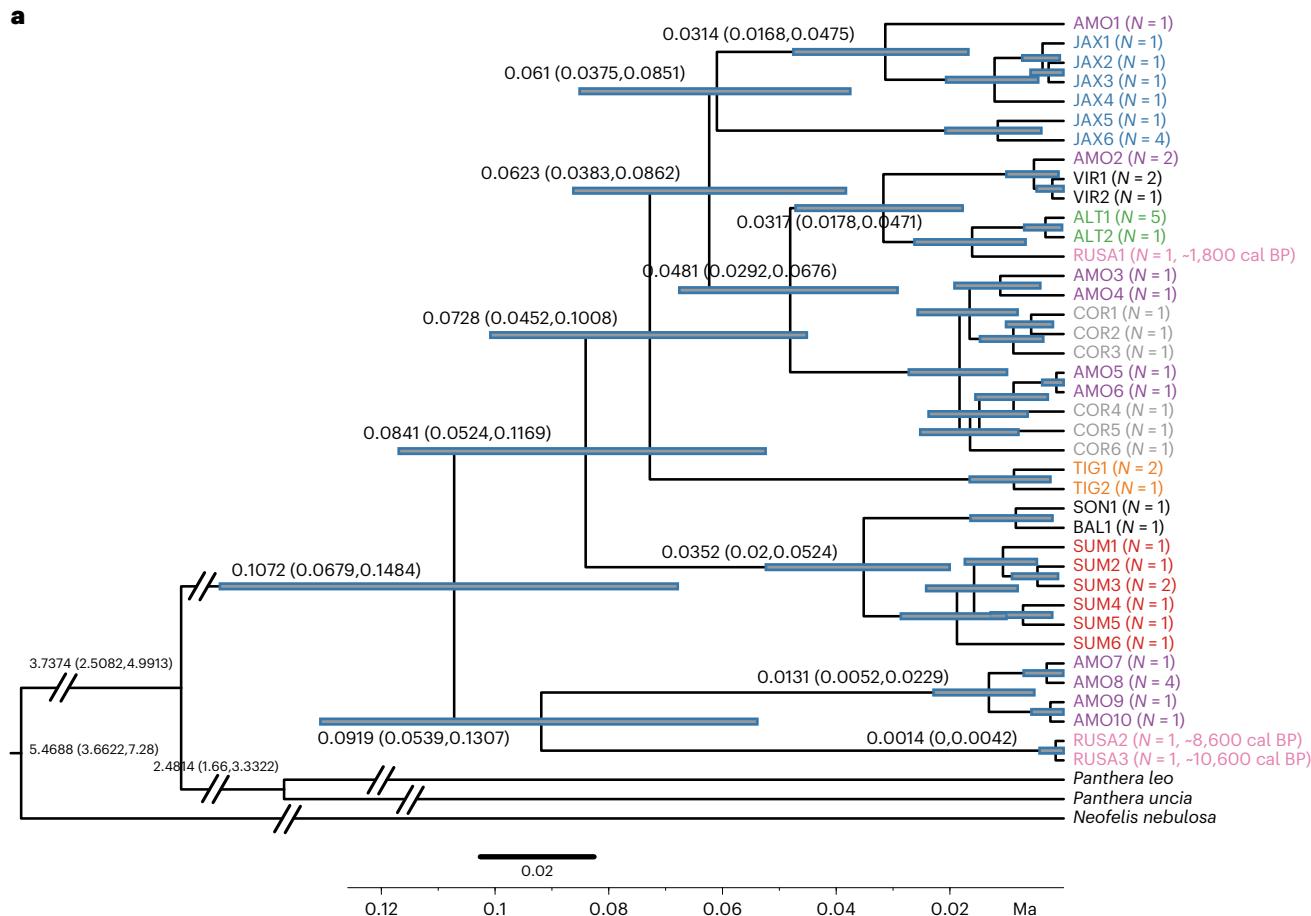
### Demographic modelling and coalescence dating

With genomic information from extinct and extant tigers, approaches of demographic modelling were performed to elucidate a most comprehensive evolutionary history of the species. We built f-statistics-based admixture graphs to integrate multiple ancient admixture scenarios revealed by the D-statistics into a single model (Fig. 4c). The best-fitting admixture graph illustrated three major branches after tigers split from the *Panthera* outgroup, including one extinct 'ghost lineage' (a in Fig. 4c) that contributed to approximately half of the genomic ancestry of the ancient Russian tiger (RUSA21), one phylogroup (b) comprising all extant tiger subspecies and one early-divergent lineage (d) eventually leading to Bengal tigers. The other half of RUSA21's genome was from genetic mixture with another ancestral lineage in Northeast Asia (c) that was associated with today's Amur and South China tigers.

The subsequent divergence leading to modern tiger subspecies can be further divided into two waves, with short relative branch lengths in 'qpGraph' that are indicative of a rapid and nearly simultaneous radiation: one lineage (e) became the ancestor to the Sumatran tiger and probably other Sundaland populations, and another lineage (f) established the other subspecies on mainland Asia including an early split (g) leading to the South China tiger. Both admixture graph modelling (Fig. 4c) and TreeMix phylogenetic approaches (from 1 to 7, Extended Data Fig. 8) supported a model where the admixed ancestry of the Caspian tiger in Central Asia derived from (1) an ancient lineage (h in Fig. 4c) related to the Bengal tiger and (2) an East Asian lineage (i in Fig. 4c) ancestral to modern Amur and South China tigers. Overall, the demographic history considering post-divergence gene flow supported that most admixture events occurred after the earliest divergence but before the latest divergences giving rise to modern subspecies. The level of gene flow between modern subspecies is highly restricted, corresponding to their population genetic distinctiveness.

Fig. 5 | Population history modelling of tigers based on modern and ancient specimens. a, Mitogenome divergence time estimated with BEAST2. Mitochondrial haplogroups are colour coded by the subspecies (modern tigers) or population (ancient Russian Far East tiger) source of the specimen, and the number of samples sharing the mtDNA haplotype are indicated in parentheses. The blue bar and the number in brackets indicated the 95% HPD intervals of the estimated divergence time. **b**, Demographic model inferred from G-PhoCS integrating the divergence time (proportional to the height of each bar), population size (proportional to the width of each bar) and total migration rate (in percentage), with the direction indicated by dashed arrows. Abbreviations for the specimens are as in Fig. 1; OUT, *P. leo* or *P. uncia* as the outgroup.

a

b

Present day

Last Glacial Maximum
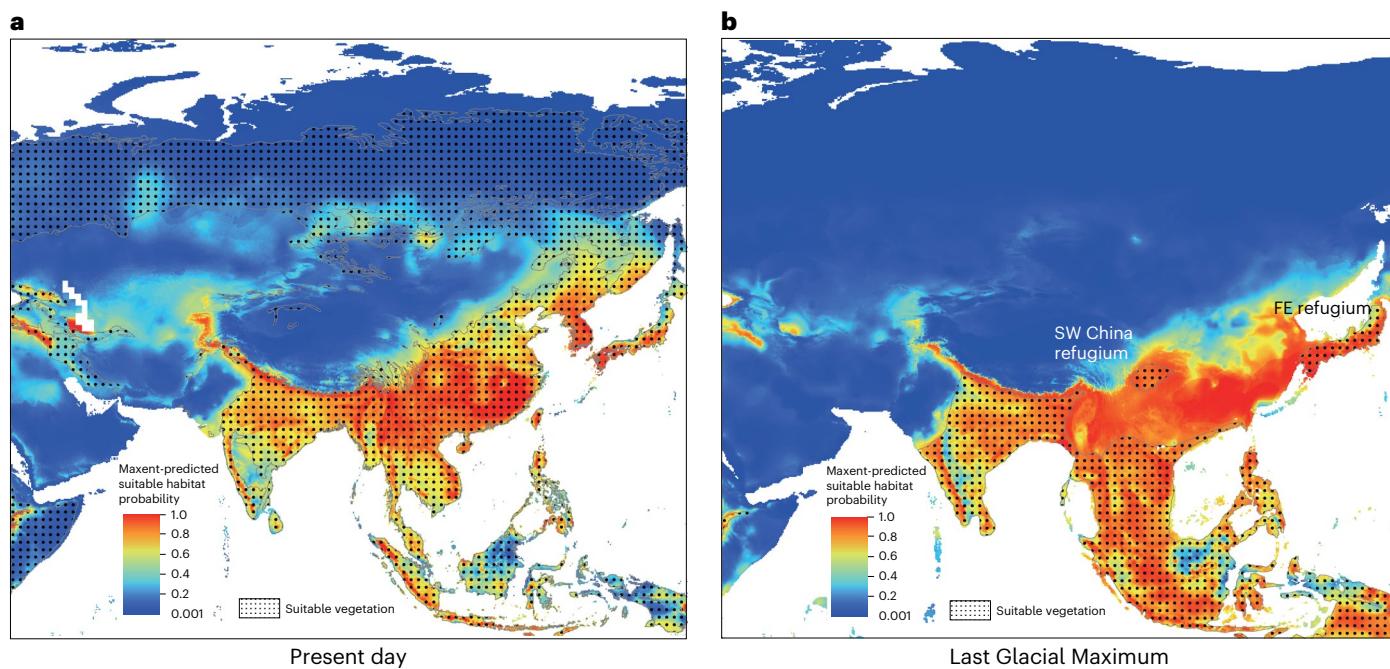
Fig. 6 | The projected tiger habitat suitability at present and during the LGM based on ENM in Maxent overlaid with the preferred vegetation layer. **a,b**, Pojected tiger habitat suitability at present (**a**) and during the LGM (**b**). The Maxent model was built with present-day tiger occurrence data collected from the literature, based on nine bioclimate variables. The probability of habitat suitability during the LGM was projected on the basis of contemporary climate variables. Possible refugia with suitable habitat for the tiger during the LGM are highlighted in Southwest China (SW China) and the Far East (FE). Different colours refer to different levels of likelihood of tiger presence. The dotted regions represent vegetation zones suitable for tigers based on data assembled from the IUCN Red List of Threatened Species for the present-day layer[88] and from published studies for the LGM map[89].

We performed coalescent dating with mitogenome data using BEAST2 (Fig. 5a). Consistent with previous studies[12–14,16,20], the TMRCA of all tigers' mitochondrial haplotypes was traced back to 107,200 yr ago (95% confidence interval (CI): 67,900–148,400 yr ago). The ancient RFE tiger from c.10,600 yr BP clustered within the modern tiger clade and hence did not alter the species' matrilineal TMRCA estimates based on modern tigers[12–14,16,20].

G-PhoCS demographic modelling based on whole-genome data (Fig. 5b) revealed a substantially more recent coalescence time of all tigers at about 19,550 yr ago (95% CI: 13,610–26,630 yr ago). The last coalescent event was the split of the Malayan tiger from the Indochinese tiger, estimated at approximately 5,740 yr ago (95% CI: 1,640–14,350 yr ago). The total post-divergence migration between subspecies was low at 3–12% and mostly between geographically adjacent populations. Although caution should be taken concerning the limited sample size, low genome sequencing coverage and data quality constraints associated with zooarchaeological specimens in complex demographic modelling, our genome-wide coalescence time estimates fall within previous estimations. The nuclear genome coalescence estimated by G-PhoCs indicates the time of the 'final clear split' between populations, while the actual divergence may begin much earlier and was possibly reflected by the mitogenome coalescence time.

Given the recent divergence time estimate, it is no surprise that the 10,600-yr-old ancient tiger specimen from the Russian Far East, the extinct Caspian tigers and the South China tigers followed a similar effective population size trajectory until 10,000 yr ago in the pairwise sequential Markovian coalescent (PSMC) model (Extended Data Fig. 9).

A historical tiger refugium in Southwest China

The taxonomy of the South China tiger is the most controversial among all, as the tiger is extinct in the wild and only survived by an inbred captive population derived from a small number of founders of uncertain origin[43]. Phylogenomic patterns of mitochondrial paraphyly and autosomal monophyly in South China tigers, as well as demographic history and gene flow inference, jointly rendered support for a distinct subspecies *P. t. amoyensis* and indicated a plausible scenario for resolving the longstanding controversy over its origin. The non-unique AMO mtDNA haplotypes (that is, shared with other subspecies) were restricted to eastern China but were absent in the west, such as Shaanxi, Chongqing and Guizhou, where only the basal lineage unique to South China tigers was present (Figs. 1 and 2b). The geographic distribution of AMO mtDNA haplotypes in China prompted a hypothesis that Southwest China may have once been a refugium harbouring a relic tiger population, while eastern China became a genetic melting pot for various ancient lineages expanding into this region only as biogeographical conditions became optimal.

To test this hypothesis, we performed ecological niche modelling (ENM) to project tiger habitat suitability under the influence of climate fluctuation from the Last Interglacial (LIG, between c.130,000 and 115,000 yr ago) to the present (Fig. 6). The climate dynamics associated with the potential tiger range roughly corresponded with alternating episodes of glacier advance and retreat. During mild and humid interglacial periods such as the LIG, Mid-Holocene or present day (Fig. 6a and Extended Data Fig. 6), suitable tiger habitat was widespread and continuous throughout much of Asia, from the Caspian Sea and the Indian subcontinent in the west, to Siberia in the north, to the Sunda Islands in the south and to the Japanese archipelago in the Far East[10,14]. During cool and dry periods such as the LGM (~22,000 yr ago, Fig. 6b), although reduced sea level may have created land bridges to facilitate dispersal, substantial climate cooling probably exerted a profoundly negative impact on tiger landscapes, resulting in local extinctions and range fragmentations.

When ecological niche models were overlaid with the vegetation types suitable for tigers (Fig. 6), a further-restricted tiger distribution during an LGM-like period was apparent (Fig. 6b), in which the majority of today's eastern part of China was composed of temperate grassland

interspersed with sparse woodland and was not optimal habitat for tigers. In contrast, two possible northern Asian refugia that may have sustained tigers and prey populations were revealed, including one pocket zone in the mountains of Southwest China and another on the Japanese archipelago. Although the arrival and extinction of tigers in ancient Japan remained somewhat elusive[44–46], the proposed refugium in Southwest China fit the scenario that tigers carrying the haplotypes basal in tiger mitochondrial phylogeny (that is, mtDNA haplotypes AMO7, AMO8, AMO9 and AMO10 in Fig. 2b) probably survived in the refugium and only presented in modern South China tigers.

Genetic admixture in eastern China during tiger expansion

Multiple genetic studies of the tiger[15,16,20] converged on a low level of genomic diversity that corresponded to a long-term small effective population size in the species. TMRCA for the mitogenome suggested that this prolonged contraction was probably associated with the beginning of the Last Glacial Period ~115,000 yr ago and subsequently differentiated into multiple phylogroups associated with various habitats and selection pressures (Fig. 5).

At the advent of a glacial period, driven by the dry and cold climate, the connection between some tiger populations across mainland Asia might have disappeared and relic lineages would have become scattered across the fragmented landscape. One ancestral lineage carrying the unique South China tiger mitochondrial haplotype probably survived in mountainous Southwest China. The long-lasting isolation ended after the glacial maximum. The post-bottleneck radiation centre was probably located in the Indochina region and started with mainland–Sundaland isolation with gene flow (M1 and G1 in Fig. 1b), followed by the split of the Bengal tiger to the Indian subcontinent (M2 in Fig. 1b). With a warmer climate and return of habitat across the Northern Hemisphere especially at the high latitudes, the ancient populations in Indochina rebounded and dispersed into eastern China (M3 in Fig. 1b), where they encountered the relic tiger population expanding out of the Southwest China refugium (M4 in Fig. 1b). The eastern part of China, then covered with habitats suitable for tigers, such as forest and shrubland, might have functioned as a 'melting pot' where distant and previously isolated lineages merged to create the genetic makeup of contemporary South China tigers.

After the last glacier completely retreated within the past 12,000 yr, tigers in eastern China continued their dispersal to the north and reached Northeast Asia, probably in multiple waves driven by climate fluctuations. One of the earlier waves probably involved an extinct lineage represented by the 10,600-yr-old tiger specimen RUSA21 and the latest turnover gave rise to the modern Amur tiger (M5 and G2 in Fig. 1b). The Caspian tiger in Central Asia originated from two sources, first established from the expansion of an ancestral Northeast Asian population through the Siberian corridor (M6 in Fig. 1b) and then admixed with populations from the Indian subcontinent (G3 in Fig. 1b). The pattern of modern tiger subspecies divergence was completed following the last wave of post-LGM migration.

Conclusions and implications to tiger conservation

In this study, analysis of tiger genomes spanning the past 10,000 yr revealed a comprehensive evolutionary history for the species. Population structure and genome-wide phylogenetic monophyly supported the South China tiger (*P. t. amoyensis*) and Caspian tiger (*P. t. virgata*) as distinct subspecies, yet cytonuclear discrepancy also revealed ancient genetic admixture from other lineages contributing to their origin. Such patterns led to the proposed scenario involving a cryptic tiger refugium in Southwest China[47,48], genetic homogenization of various distinct lineages in eastern China following the return of habitat, and multiple waves of population expansion, admixture and turnover in the RFE area.

Ancient genomes showed that during the process leading to the phylogeographic partitioning of modern tiger subspecies, Southwest China might have served as a Late Pleistocene refugium for relic tiger lineages and eastern China as an ancient genetic melting pot for various lineages to meet and merge, as they expanded their range following the return of suitable habitats. The expansion of tigers across China en route to Northeast Asia and then westbound to Central Asia connected the northern subspecies, such as the Amur and Caspian tigers, with the southern mainland subspecies, such as the Bengal and Indochinese tigers. China is the only country in the world that once harboured five out of the nine modern tiger subspecies, or all those from mainland Asia[49]. These new findings further highlight the significance of mainland China as a stepping stone in the evolutionary history of tigers.

A thorough understanding of the tiger's evolutionary genomic diversity is crucial for conservation and recovery planning for this charismatic megafauna. Management strategies for the tiger, both in the wild and in captivity, have been influenced by perceptions of its geographical variation and subspecific taxonomy. With genomic data retrieved from nowadays extinct populations, this study resolved the basal rooting in the tiger phylogeny and illuminated cryptic biogeographic refugia and dispersal paths critical for the formation of extant landscape genetic patterns. These results promise to contribute to the much-needed scientific foundation guiding conservation, such as restoring tigers to extinct landscapes or regenerating historical connectivity across fragmented habitats.

Methods

Ancient specimens and radiocarbon dating

Twenty-five ancient felid skeletal samples were collected during palaeontological excavations by the Far Eastern Branch of the Russian Academy of Sciences from limestone caves in the Primorsky Territory and the Jewish Autonomous Region, Russian Far East: Yunyh Speleologov Cave (43° 29′ N, 132° 26′ E), Letuchaya Mysh Cave (42° 59′ N, 133° 05′ E), Tetukhinskaya Cave (44° 35′ N, 135° 36′ E) and Koridornaya Cave (48° 00′ N, 130° 59′ E) (Extended Data Table 1 and Fig. 1, and Supplementary Table 2). These specimens were identified to be from big cats on the basis of morphological evaluation. Seven samples yielding genome sequencing data were selected for accelerator mass spectrometry (AMS)-^{14}C radiocarbon dating (Supplementary Table 1) at Peking University and Beta Analytic Co. Calibration was performed on the basis of the IntCal13 atmospheric curve[49] with OxCal (v.4.2.4)[50,51].

Historical samples, such as bones, pelts and teeth, were assembled from 31 South China tigers located in various museums, institutes or private collections in China, Europe and the United States (Extended Data Table 1 and Fig. 1, and Supplementary Table 2). All specimens were verified as having geographic origins in the traditionally recognized *P. t. amoyensis* (AMO) range, covering ten provinces, namely, Hubei, Hunan, Shaanxi, Yunnan, Guangdong, Jiangxi, Fujian, Shanxi, Sichuan and Guizhou. The five nomenclatural type specimens of *P. t. amoyensis* (type locality: Hankou, China)[34], including one holotype (museum ID: 3311) and four paratypes (3305, 3306, 3307, 3308) were sampled from Musée Zoologique de la Ville de Strasbourg, France.

Six Caspian tiger specimens (VIR) were chosen for whole-genome sequencing from a previous museum collection[11] with partial mtDNA sequencing information. These samples represented the key *P. t. virgata* distribution in Uzbekistan, Kazakhstan, Turkmenistan, Tajikistan and Azerbaijan (Extended Data Table 1 and Fig. 1, and Supplementary Table 2).

DNA extraction, library preparation and capture enrichment

We processed degraded specimens in dedicated ancient DNA facilities at the School of Life Sciences, Peking University, China, and the Globe Institute, University of Copenhagen, Denmark. For DNA extraction, we evaluated and chose from several optimized ancient DNA protocols on the basis of sample type. Bone samples were digested in proteinase K buffer[52]. Tissue samples were digested with Buffer ATL from the DNeasy blood and tissue kit (Qiagen) with proteinase K.

For both bone and tissue samples, pre-digestion with a small volume of digestion buffer at 37 °C was applied to remove external DNA contamination[52,53], followed by a phenol-chloroform extraction protocol. The digest was then applied to a MinElute spin column (Qiagen) using a 10-fold volume of binding buffer[53] and further processed following manufacturer protocol.

Next-generation sequencing libraries were prepared with various methods optimized for ancient DNA. For DNA from relatively fresh samples, libraries were constructed with the NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs). DNA libraries for museum or archaeological samples were built using the blunt-end single-tube method[54]. Libraries were sequenced on an Illumina HiSeq 2500 or X Ten platform at BIOPIC, Peking University or Novogene. A 'screen and boost' sequencing strategy was used, in which 3–5 Gb of data from each library was retrieved, followed by subsequent boost sequencing according to the endogenous DNA content from the first round. For ancient RFE and Caspian tiger specimens, a uracil-specific excision reagent (USER, a mixture of uracil DNA glycosylase and the DNA glycosylase-lyase Endonuclease VIII) treatment was applied to the same batch of DNA used for library construction during the second round of boost sequencing.

For ancient RFE tiger samples, six samples with an endogenous DNA content above 0.5% were chosen for mitogenome capture enrichment. We used the tiger 'myBaits Expert Mito' kit (Daicel Arbor Bioscience) following manufacturer protocol (v.4.01), with the hybridization temperature set at 60 °C. The DNA used for mitogenome capture was pre-treated with USER enzyme mix.

PCR amplification and Sanger sequencing of mtDNA fragments were performed for all historical South China tiger specimens following previously described procedures[12]. Partial mtDNA haplotypes and subspecies-diagnostic sites were compared to tiger voucher specimens[12,16,20] to infer the maternal genetic ancestry of the individuals and to screen for samples suitable for further whole-genome resequencing.

Data processing
Quality control and alignment. Two separate data processing pipelines were implemented for the historical specimens (~100 yr old) from the Caspian and South China tigers or the ancient specimens (1,000–10,000 yr old) from the Russian Far East. We used Cutadapter (v.2.1)[55] to trim sequencing adapters. Then, 2 BP for recent samples or 5 BP for ancient samples were removed from both ends of the reads. Bases with a minimum quality of 30 at both ends and a minimum read length of 25 BP were kept for subsequent mapping. The retained reads were mapped to the tiger reference genome[13]. For historical specimens, we used the BWA-MEM[56] algorithm to map the data. For ancient specimens, paired-end reads were collapsed with AdapterRemoval (v.2.2.2)[57], allowing a minimum length of 30 BP, and a BWA-backtrack algorithm was used for mapping, with the option 'seeding' disabled. The mapped paired-end reads were treated as two separate single-end reads during the removal of PCR duplicates. SAMtools (v.1.9)[58] was used to retain reads with a minimum mapping quality of 30 and remove PCR duplicates.

Authentication of ancient DNA data
DNA damage patterns characteristic of ancient DNA were examined using mapDamage2[59] for the Russian archaeological specimens. All seven specimens with high-throughput sequencing data showed increased C to T and G to A substitutions at the 5′ and 3′ read ends. Pre- and post-USER mix treatment data for RUSA21 were compared (Extended Data Fig. 1), and near-complete removal of the damage pattern was observed, suggesting the presence of ancient DNA. For ancient specimens, only libraries with USER mix treatment were included in the final dataset.

Sample quality was evaluated by comparing the relative error rate to a high-quality sample (assumed error free) and an outgroup[60] using ANGSD[61] (Extended Data Fig. 2). Sample RFET0002 was used as the error-free individual and the reference genome was used as the outgroup individual to denote the ancestral state. The relative error rates measure the excess of derived alleles of a tested sample compared to the error-free individual.

Genotype calling
Single-nucleotide polymorphisms (SNPs) were identified using GATK v.3.8 HaplotypeCaller and GATK v.4.0 GenotypeGVCFs[62–64] on the basis of our dataset including 32 tigers from a previous study[20]. A hard-filtering procedure was performed using BCFtools (v.1.9)[65] and homemade scripts on the basis of the following criteria: (1) only biallelic SNPs were retained and SNPs within 5 BP of an indel were excluded; (2) BaseQRankSum (−1.96 < BQRS < 1.96) and MappingQualityRankSum (MQRS > −1.96) filters were applied; (3) genotype GT was marked as missing for loci in samples with a sequencing depth more than 4-fold of the average sample sequencing depth; (4) for rare alleles found only in degraded specimens, a minimum allele count of three and presence in at least three different specimens were required; (5) SNPs with a missing rate above 50% were excluded; and (6) only SNPs in scaffolds with a minimum length of 10 kb were included. The hard-filtered dataset was further filtered according to the needs of various downstream analyses and after considering variation in ancient DNA damage and sequencing depth among specimens. The distribution of relative mutation rates by mutation type (Extended Data Fig. 2) was summarized across all individuals on the basis of the number of a specific type of mutation, normalized by the total number of SNPs retained in each tiger.

Autosomal neutral variant identification
We assembled autosomal regions by identifying and excluding the putative scaffolds related to the X and Y chromosomes (Supplementary Fig. 1). Sex chromosome-derived scaffolds were identified by comparing the sequencing depth ratio between female and male tigers and the homologous scaffolds with the cat (*Felis catus*) sex chromosomes. In the first approach, only scaffolds with a minimum length of 10 kb were assessed. We identified scaffolds with a relative average sequencing depth in females more than 1.5-fold of that in males as derived from the X chromosome, and those less than 0.2× of that in males as from the Y chromosome. Alternatively, we first applied a sliding-window method to generate reads from the tiger reference genome[13], with a window size of 100 BP and a step size of 50 BP. The reads from each scaffold were then mapped to the domestic cat sex chromosomes (Felis_catus_9.0, RefSeq assembly accession: GCF_000181335.3) using BWA v.0.7.12, and a scaffold with mapping rate over 10% was considered sex-chromosome related. The results from the two methods were combined to obtain a total length of 111,848,290 BP (130,557,009 BP in *F. catus*) of X chromosome-related scaffolds and 1,481,767 BP (1,855,781 BP in *F. catus*) of Y chromosome-related scaffolds. The neutral autosomal regions were gathered after filtering out the coding regions and 1 kb upstream and downstream of a gene from the recognized autosomal scaffolds.

Mitochondrial DNA phylogeny
Mitochondrial DNA sequences were assembled independently, with caution taken to exclude nuclear mitochondrial sequence (*numt*) contaminants[66–68]. After passing quality control, the raw reads as well as those from mitogenome capture enrichment were mapped to a previously confirmed cytoplasmic, *numt*-free mitogenome reference (NCBI ID: KP202268). The mapped reads were then processed in Geneious (v.9.1.5)[69] as input for iterative de novo assembly to further exclude *numt* contaminants. For the only sample with a relatively low sequencing depth (RUSA23, mitogenome coverage lower than 10×), we manually inspected and corrected the consensus sequence for *numt* contamination. The highly variable control

region was excluded from the analysis due to its unstable assembly results with too many ambiguous sites. We obtained three different mitogenomes from ancient RFE tigers, ten from South China tigers and two from Caspian tigers.

We employed four different methods[20] to reconstruct the tiger mitogenome phylogeny, including neighbour joining (NJ), maximum parsimony (MP), maximum likelihood (ML) using PAUP (v.4.0b10)[70] and Bayesian methods using MrBayes (v.3.2.6)[71]. Two *Panthera* species, the lion (*P. leo*, NCBI: KP202262.1) and snow leopard (*P. uncia*, NCBI: KP202269.1), and the clouded leopard (*Neofelis nebulosi*, NCBI: KP202291.1), the sister taxa of *Panthera*, were used as outgroups. TIM3 + G + I was chosen as the optimal nucleotide substitution model by jModelTest (v.2.1.4)[72]. The statistical support for the NJ, MP and ML tree topologies was assessed in PAUP on the basis of 1,000 bootstraps. Bayesian inference was performed with two independent Markov chain Monte Carlo (MCMC) runs, each set for 10,000,000 generations. Tree sampling was carried out every 1,000 generations and the first 25% of the iterations was discarded as burn-in.

Whole-genome phylogeny
The whole-genome phylogeny was inferred on the basis of neutral autosomal SNPs, including the variants present only in outgroups. Unless otherwise specified, this and following analyses were restricted to transversion sites in autosomal neutral regions to minimize the effect of ancient DNA damage. The dataset consisted of 3,838,774 sites that were concatenated into one consensus sequence for each sample. The phylogeny was then reconstructed with RAxML (v.8.2.11)[73] using the GTAGAMMA model and node support was evaluated with 100 bootstrap replicates.

Population genetic structure analysis
We performed PCA with 'smartpca' in EIGENSOFT (v.6.1.4)[74] and calculated PCs using all tiger samples with an average sequencing depth higher than 1×. The final dataset consisted of 1,242,142 SNPs using transversion sites only. The ancient RFE tiger specimen RUSA21 and other low-coverage South China tiger specimens were then projected using the 'lsqproject' method implemented in 'smartpca'. We also calculated pairwise genetic differences between populations (F_{ST}) using 'smartpca' in EIGENSOFT (v.6.1.4)[74] (Supplementary Table 3).

We inferred population genetic structure with ADMIXTURE (v.1.3.0)[75] on the basis of the same SNP panel used in PCA. The number of assumed ancestral populations (*K*) was set from 2 to 10 (Supplementary Fig. 3). For each *K*, we ran 10 independent replicates with different starting seeds and chose the replicate with the highest likelihood.

Genotype likelihoods-based population structure analyses
We performed population structure analyses on the basis of a genotype likelihoods dataset estimated with ANGSD (v.0.940)[61]. Specifically, we used the GATK model (-GL 2) and included reads with mapping quality above 30 (-minMapQ 30) and bases in the reads with sequencing quality above 20 (-minQ 20). Variable sites with a minor allele frequency of less than 0.02, or with more than 50% of the samples having missing genotype information were removed. Transitions were also removed to minimize the effect of ancient DNA-induced errors.

Population structure was explored with the genotype likelihoods dataset using PCAngsd (v.1.11)[76] for PCA and NGSadmix (v.32)[77] for supervised genetic component modelling. To explore the imbalanced sample size per population, we tested various numbers of maximum samples per population and randomly selected samples to perform PCA, and projected the rest of the samples using Procrustes projection with the R package 'vegan' (Supplementary Fig. 2).

Heterozygosity (Extended Data Fig. 3) was also estimated using genotype likelihood-based method in ANGSD. Only transversions were included for this analysis, and samples with sequencing depth lower than 4× were excluded.

TreeMix phylogeny of population relationships
We used TreeMix (v.1.13)[78] to estimate the historical relationships among populations with admixture and infer possible gene flow events on the basis of allele frequency in each population. The same panel of autosomal variants used for phylogeny reconstruction was applied. To account for the effect of linkage disequilibrium, we set the -*k* flag at 1,000 to group nearby SNPs into blocks of 1,000 consecutive SNPs. Two historical South China tiger specimens (NGHE0002 and YANG0014), which displayed admixed genetic ancestry based on ADMIXTURE analysis, were designated an independent population, 'AMO_MIX'.

TreeMix runs were performed with the migration bands (-m) set from 0 to 7. Each migration band was run with 100 independent replicates using different starting seeds and the replicate with the highest likelihood is presented.

D-statistic estimation of gene flow
We applied *D*-statistics to infer possible gene flow between different tiger populations, specifically testing for excessive allele sharing as implemented in 'qpDstat' of ADMIXTOOLS (v5.1)[79]. The analysis was conducted on the basis of the same dataset used for whole-genome phylogeny reconstruction and the following scenarios were tested: (1) the possible post-divergence gene flow between the Caspian (VIR) and Bengal tigers (TIG), or *D*(Outgroup, TIG, VIR, other populations (PopX)); (2) the asymmetric relationship among the ancient RFE tiger (RUSA), its closely related populations (the South China tiger AMO and Amur tiger ALT) and other populations, or *D*(Outgroup, other populations (PopX), AMO/ALT, RUSA); (3) the asymmetric relationship of one high-quality South China tiger genome (sample ID: HPS) to other South China tigers, or *D*(Outgroup, other populations (PopX), AMO (HPS), all other South China tigers (PopX)); and (4) the possible post-divergence gene flow between the Sumatran tiger (SUM) and other tigers, or *D*(Outgroup, SUM, TIG/COR/JAX (Pop3), all other populations (PopX)). The lion *P. leo* (PLE) and snow leopard *P. uncia* (PUN) were used as outgroup taxa.

Admixture graph modelling with 'qpGraph'
We built admixture graphs using 'qpGraph'[79], which is based on *f*-statistics and takes into account genetic drift and mixture proportions, to model the evolutionary pathways and relationships among tiger populations. The same dataset including both modern and ancient tigers in *D*-statistics was included. We picked one sample with the highest sequencing depth to represent each population and searched the graph space using a heuristic algorithm implemented in 'qpBrute'[80,81]. A base graph was first constructed with two outgroups and the Sumatran tiger, followed by addition of the remaining populations in two rounds. The first round included three modern subspecies (JAX, TIG and ALT) and one ancient population (RUSA) and resulted in 6,727 uniquely fitted graphs. The top three graphs with the lowest *z*-scores, the highest likelihoods and no zero-length branches were set as the starting graph in the second batch, to which the remaining three subspecies (COR, AMO and VIR) were added. Only one graph from the 15,179 unique fits without outliers or zero-length branches is presented as the best-fitting admixture graph model.

Demographic inference with PSMC
We applied the PSMC[82] model to infer the historical dynamics of effective population size of various tiger subspecies. A consensus sequence was generated from the individual with the highest sequencing depth to represent each modern or ancient tiger population. Only the autosomal neutral region was included and the rest of the genome was masked. The parameters for PSMC modelling were set as -N25 -t20 -p '4 + 25*2 + 4 + 6' and 100 bootstrap replicates were run to assess the robustness of the results. A mutation rate of 6.4×10^{-9} substitutions per site per generation was recalculated[20] on the basis of the autosomal neutral SNPs from this study, a generation time of 5 yr for the tiger, and

the overall divergence between the tiger and other *Panthera* species such as the lion and snow leopard. To minimize the effect of medium sequencing coverage, we empirically corrected for underestimation of heterozygosity in our ancient samples by setting '-M' in the model and assuming similar effective population size for each tiger population around 1 Ma.

Demographic inference with G-PhoCS

The demographic history of tigers, including the historical effective population sizes, divergence times and gene flow scenarios, was modelled in G-PhoCS[83] on the basis of the phylogeny inferred by TreeMix (Extended Data Fig. 8 and Supplementary Fig. 4). Two individuals with the highest sequencing depth were chosen to represent each modern tiger subspecies and one sample was from the ancient RFE population. The same dataset used for PSMC modelling was applied and 44,092 autosomal loci were identified, each spanning l kb with a minimum inter-locus gap of 50 kb and a maximum missing rate of 5%. Four alternative combinations of individual and different outgroup species were evaluated to ensure model convergence. We first ran the model with all possible migration bands between different branches using 5,000 randomly chosen loci in two independent replicates. All migration bands were split into six groups to reduce the total number of parameters in each model. The total migration rate for each migration band was calculated as described in the G-PhoCS manual and migration bands with a total migration rate above 0.1 in at least one running batch were kept as the final model. We then ran the final model (Supplementary Fig. 4) in two parallel replicates for each scenario of individual combinations. Estimates of model parameters (Supplementary Fig. 5) were cross-checked for various runs and combined to obtain the final results. The divergence times for *P. tigris* and *P. leo* at 3.72 Ma (95% CI: 2.04–7.60 Ma), and *P. tigris* and *P. uncia* at 2.67 Ma (95% CI: 1.14–5.92 Ma) were used as calibration values for the model[2]. A generation time of 5 yr for the tiger was applied.

Mitogenome coalescence time estimation

The coalescence time of modern and ancient tiger mitogenomes was estimated in BEAST (v.2.5.2)[84] on the basis of the same nucleotide substitution model (TIM3 + I + G, used in phylogeny inference), a strict-clock setting and a coalescent-exponential-population model. The coalescence times between *N. nebulosa* and *Panthera* spp. (6.37 Ma, 95% CI: 4.47–9.32 Ma), and between *P. leo* and *P. uncia*/*P. tigris* (3.72 Ma, 95% CI: 2.04–7.60 Ma) were used as calibration values. MCMC was performed with five independent runs for 50,000,000 iterations each. Sampling was performed every 1,000 generations and the first 20% of values was discarded as burn-in.

Ecological niche modelling

We built ecological niche models in Maxent (v.3.4.0)[85] to correlate presence data with the species' ecological requirements and to project tiger habitat suitability during different geological periods. A total of 781 tiger locality points (Supplementary Fig. 6) were assembled from published studies[10] and records from the IUCN Cat Specialist Group[86,87]. Bioclimate variables from WorldClim (1.4)[88] were used as predictors and 8 of the 19 variables (Supplementary Table 4) were retained after removing the highly correlated ones ($|r| \geq 0.8$). The models were built at a resolution of 10 arc-min. The spatial extent ranged from 20° W to 144° E and 10° S to 82° N, and was restricted to the region with tiger occurrence to reduce bias in model fitting. The models were constructed under the modern climate with an area under the curve of 0.835 and then projected to three types, namely, the LIG period (c.120,000–140,000 yr BP), the LGM (c.22,000 yr BP) and the Mid-Holocene (c.6,000 yr BP).

Two additional suitable habitat layers for tigers were generated based on the published biome and vegetation layers for the modern[89] and LGM[90] geological periods. The binary suitability (suitability vs unsuitability) of different biomes and vegetation types for tigers (Supplementary Table 5) was manually designated according to tigers' preferred habitat classification as listed in the IUCN Red List of Threatened Species[17].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The next-generation-sequencing raw data of the tiger samples have been deposited in the Sequence Read Archive (BioProject ID: PRJNA822019). The data processing pipeline is available at https://github.com/xinsun1/Ancient_tiger_pop_gen. The processed data file is available at https://doi.org/10.5061/dryad.73n5tb324.

References

1. Johnson, W. E. et al. The late Miocene radiation of modern felidae: a genetic assessment. *Science* **311**, 73–77 (2006).
2. Li, G., Davis, B. W., Eizirik, E. & Murphy, W. J. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* **26**, 1–11 (2016).
3. Hemmer, H. Fossil history of living felidae. *Carnivore* **2**, 58–61 (1979).
4. Kitchener, A. C. & Yamaguchi, N. In *Tigers of the World* (eds Tilson, R. & Nyhus, P. J.) 53–84 (Elsevier, 2010).
5. Hemmer, H. in *Tigers of the World. The Biology, Biopolitics, Management, and Conservation of an Endangered Species* (eds Tilson, R. L. & Seal, U. S.) 28–35 (Noyes Publications, 1987).
6. Mazák, J. H., Christiansen, P. & Kitchener, A. C. Oldest known pantherine skull and evolution of the tiger. *PLoS ONE* **6**, e25483 (2011).
7. Werdelin, L., Yamaguchi, N., Johnson, E. & O'Brien, S. J. In *Biology and Conservation of Wild Felids* (eds Macdonald D. & Loveridge A.) 59–82 (Oxford Univ. Press, 2010).
8. Mazak, V. *Panthera tigris*. *Mamm. Species* https://doi.org/10.2307/3504004 (1981).
9. Luo, S.-J., Liu, Y.-C. & Xu, X. Tigers of the world: genomics and conservation. *Annu. Rev. Anim. Biosci.* **7**, 521–548 (2019).
10. Cooper, D. M. et al. Predicted Pleistocene–Holocene range shifts of the tiger (*Panthera tigris*). *Divers. Distrib.* **22**, 1199–1211 (2016).
11. Driscoll, C. A. et al. Mitochondrial phylogeography illuminates the origin of the extinct Caspian tiger and its relationship to the Amur tiger. *PLoS ONE* **4**, e4125 (2009).
12. Xue, H.-R. et al. Genetic ancestry of the extinct Javan and Bali tigers. *J. Hered.* **106**, 247–257 (2015).
13. Cho, Y. S. et al. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat. Commun.* **4**, 2433 (2013).
14. Wilting, A. et al. Planning tiger recovery: understanding intraspecific variation for effective conservation. *Sci. Adv.* **1**, e1400175 (2015).
15. Armstrong, E. E. et al. Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. *Mol. Biol. Evol.* **38**, 2366–2379 (2021).
16. Luo, S.-J. et al. Phylogeography and genetic ancestry of tigers (*Panthera tigris*). *PLoS Biol.* **2**, e442 (2004).
17. Goodrich, J. et al. *Panthera tigris*. *The IUCN Red List of Threatened Species 2022* (IUCN, 2022).
18. Tilson, R., Defu, H., Muntifering, J. & Nyhus, P. J. Dramatic decline of wild South China tigers *Panthera tigris amoyensis*: field survey of priority tiger reserves. *Oryx* **38**, 40–47 (2004).
19. Tilson, R., Traylor-Holzer, K. & Jiang, Q. M. The decline and impending extinction of the South China tiger. *Oryx* **31**, 243–252 (1997).

20. Liu, Y.-C. et al. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Curr. Biol.* **28**, 3840–3849.e6 (2018).

21. Mazák, J. H. Craniometric variation in the tiger (*Panthera tigris*): implications for patterns of diversity, taxonomy and conservation. *Mamm. Biol.* **75**, 45–68 (2010).

22. van der Valk, T. et al. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).

23. Skoglund, P. & Mathieson, I. Ancient genomics of modern humans: the first decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018).

24. Orlando, L. & Cooper, A. Using ancient DNA to understand evolutionary and ecological processes. *Annu. Rev. Ecol. Evol. Syst.* **45**, 573–598 (2014).

25. Barnett, R. et al. Genomic adaptations and evolutionary history of the extinct scimitar-toothed cat, *Homotherium latidens*. *Curr. Biol.* **30**, 5018–5025.e5 (2020).

26. Paijmans, J. L. A. et al. Evolutionary history of saber-toothed cats based on ancient mitogenomics. *Curr. Biol.* **27**, 3330–3336.e5 (2017).

27. Westbury, M. V. et al. A genomic exploration of the early evolution of extant cats and their sabre-toothed relatives. *Open Res. Eur.* **1**, 25 (2021).

28. de Manuel, M. et al. The evolutionary history of extinct and living lions. *Proc. Natl Acad. Sci. USA* **117**, 10927–10934 (2020).

29. Barnett, R. et al. Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Mol. Ecol.* **18**, 1668–1677 (2009).

30. Barnett, R. et al. Mitogenomics of the extinct cave lion, *Panthera spelaea* (Goldfuss, 1810), resolve its position within the *Panthera* cats. *Open Quat.* **2**, 4 (2016).

31. Salis, A. T. et al. Lions and brown bears colonized North America in multiple synchronous waves of dispersal across the Bering Land Bridge. *Mol. Ecol.* **31**, 6407–6421 (2022).

32. Paijmans, J. L. A. et al. Historical biogeography of the leopard (*Panthera pardus*) and its extinct Eurasian populations. *BMC Evol. Biol.* **18**, 156 (2018).

33. Hu, J. et al. An extinct and deeply divergent tiger lineage from northeastern China recognized through palaeogenomics. *Proc. R. Soc. B* **289**, 20220617 (2022).

34. Hilzheimer, H. Uber einige Tigerschadel aus der Strassburger Zoologischen Sammlung. *Zool. Anz.* **28**, 594–599 (1905).

35. Zhang, W. et al. Sorting out the genetic background of the last surviving South China tigers. *J. Hered.* **110**, 641–650 (2019).

36. Baryshnikov, G. F. Late Pleistocene Felidae remains (Mammalia, Carnivora) from Geographical Society Cave in the Russian Far East. *Proc. Zool. Inst. RAS* **320**, 84–120 (2016).

37. Tiunov, M. P. & Gimranov, D. O. The first fossil *Petaurista* (Mammalia: Sciuridae) from the Russian Far East and its paleogeographic significance. *Palaeoworld* **29**, 176–181 (2020).

38. Tiunov, M. P. & Gusev, A. E. A new extinct ochotonid genus from the late Pleistocene of the Russian Far East. *Palaeoworld* **30**, 562–572 (2021).

39. Tiunov, M. P., Golenishchev, F. N. & Voyta, L. L. The first finding of Mimomys in the Russian Far East. *Acta Palaeontol. Pol.* **61**, 205–210 (2016).

40. Voyta, L. L., Omelko, V. E., Tiunov, M. P. & Vinokurova, M. A. When beremendiin shrews disappeared in East Asia, or how we can estimate fossil redeposition. *Hist. Biol.* **33**, 2656–2667 (2021).

41. Nowell, K. & Peter Jackson *Wild Cats: Status Survey and Conservation Action Plan* (IUCN, 1996).

42. Gour, D. S. et al. Philopatry and dispersal patterns in tiger (*Panthera tigris*). *PLoS ONE* **8**, e66956 (2013).

43. Luo, S.-J. et al. Proceedings in phylogeography and genetic ancestry of tigers (*Panthera tigris*) in China and across their range. *Zool. Res.* **27**, 441–448 (2006).

44. Hasegawa, Y., Takakuwa, Y., Nenoki, K. & Kimura, T. Fossil tiger from limestone mine of Tsukumi City, Oita Prefecture, Kyushu Island, Japan. *Bull. Gunma Mus. Nat. Hist.* **23**, 1–11 (2019).

45. Kawamura, Y., Kamei, T. & Taruno, H. Middle and Late Pleistocene mammalian faunas in Japan. *Quat. Res.* **28**, 317–326 (1989).

46. Hasegawa, Y. Summary of quaternary carnivore in Japan. *Mamm. Sci.* **38**, 23–28 (1979).

47. Tang, C. Q. et al. Identifying long-term stable refugia for relict plant species in East Asia. *Nat. Commun.* **9**, 4488 (2018).

48. Song, W. et al. Multiple refugia from penultimate glaciations in East Asia demonstrated by phylogeography and ecological modelling of an insect pest. *BMC Evol. Biol.* **18**, 152 (2018).

49. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

50. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720–730 (2013).

51. Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).

52. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).

53. Rasmussen, M. et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).

54. Carøe, C. et al. Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* **9**, 410–419 (2018).

55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* https://doi.org/10.14806/ej.17.1.200 (2011).

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

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

58. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

60. Orlando, L. et al. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).

61. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

62. Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

63. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

64. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

65. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

66. Antunes, A., Pontius, J., Ramos, M. J., O'Brien, S. J. & Johnson, W. E. Mitochondrial introgressions into the nuclear genome of the domestic cat. *J. Hered.* **98**, 414–420 (2007).

67. Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **39**, 174–190 (1994).

68. Kim, J.-H. et al. Evolutionary analysis of a large mtDNA translocation (*numt*) into the nuclear genome of the *Panthera* genus species. *Gene* **366**, 292–302 (2006).

69. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

70. Wilgenbusch, J. C. & Swofford, D. Inferring evolutionary trees with PAUP. *Curr. Protoc. Bioinformatics* **00**, 6.4.1–6.4.28 (2003).

71. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

72. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

73. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

74. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

75. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

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

77. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

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

79. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

80. Ní Leathlobhair, M. et al. The evolutionary history of dogs in the Americas. *Science* **361**, 81–85 (2018).

81. Liu, L. et al. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat. Commun.* **10**, 1992 (2019).

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

83. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034 (2011).

84. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

85. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. Opening the black box: an open-source release of Maxent. *Ecography* **40**, 887–893 (2017).

86. Luo, S.-J. The status of the tiger in China. *Cat. N. Spec. Issue* **5**, 10–13 (2010).

87. Smith, A. T. et al. *A Guide to the Mammals of China* (Princeton Univ. Press, 2008).

88. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

89. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51**, 933–938 (2001).

90. Ray, N. & Adams, J. M. A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000–15,000 BP). *Internet Archaeol.* https://doi.org/10.11141/ia.11.2 (2001).

Author contributions

The project was conceived and designed by S.-J.L. Laboratory work was done by X.S., Y.-C.L., Y.Z., Y.H., Y.-H.P, C.L., Y.P., M.S.V. and Y.-Y.H. Data analysis was performed by X.S., S.G. and X.-H.W. Samples were provided by M.P.T., D.O.G., C.A.D., R.-Z.Y, B.-G.L., K.J., X.X., O.U. and N.Y. The initial manuscript draft was written by X.S. and S.-J.L., with helpful input from M.T.P.G., S.J.O. and N.Y. All authors contributed to interpreting the data and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-023-02185-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-023-02185-8.

Correspondence and requests for materials should be addressed to Stephen J. O'Brien, Nobuyuki Yamaguchi or Shu-Jin Luo.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.
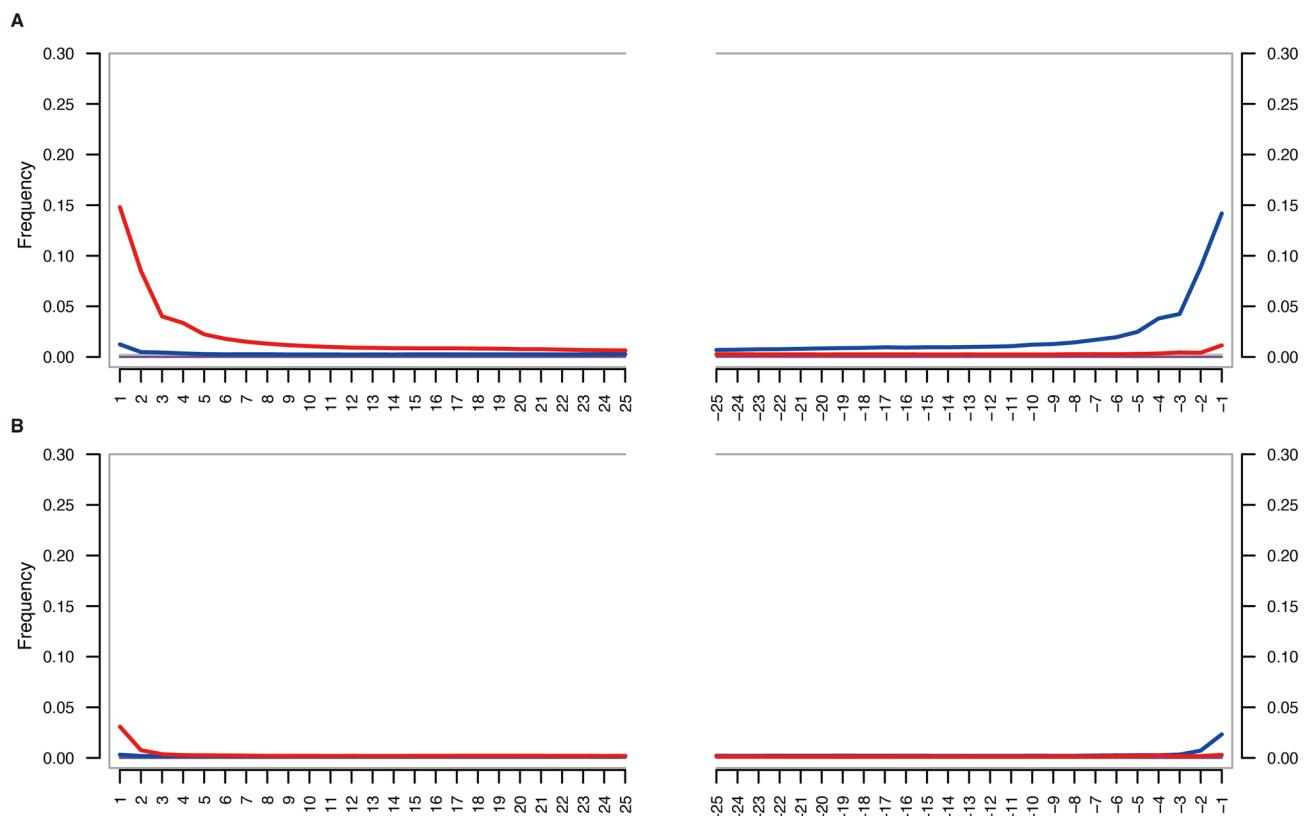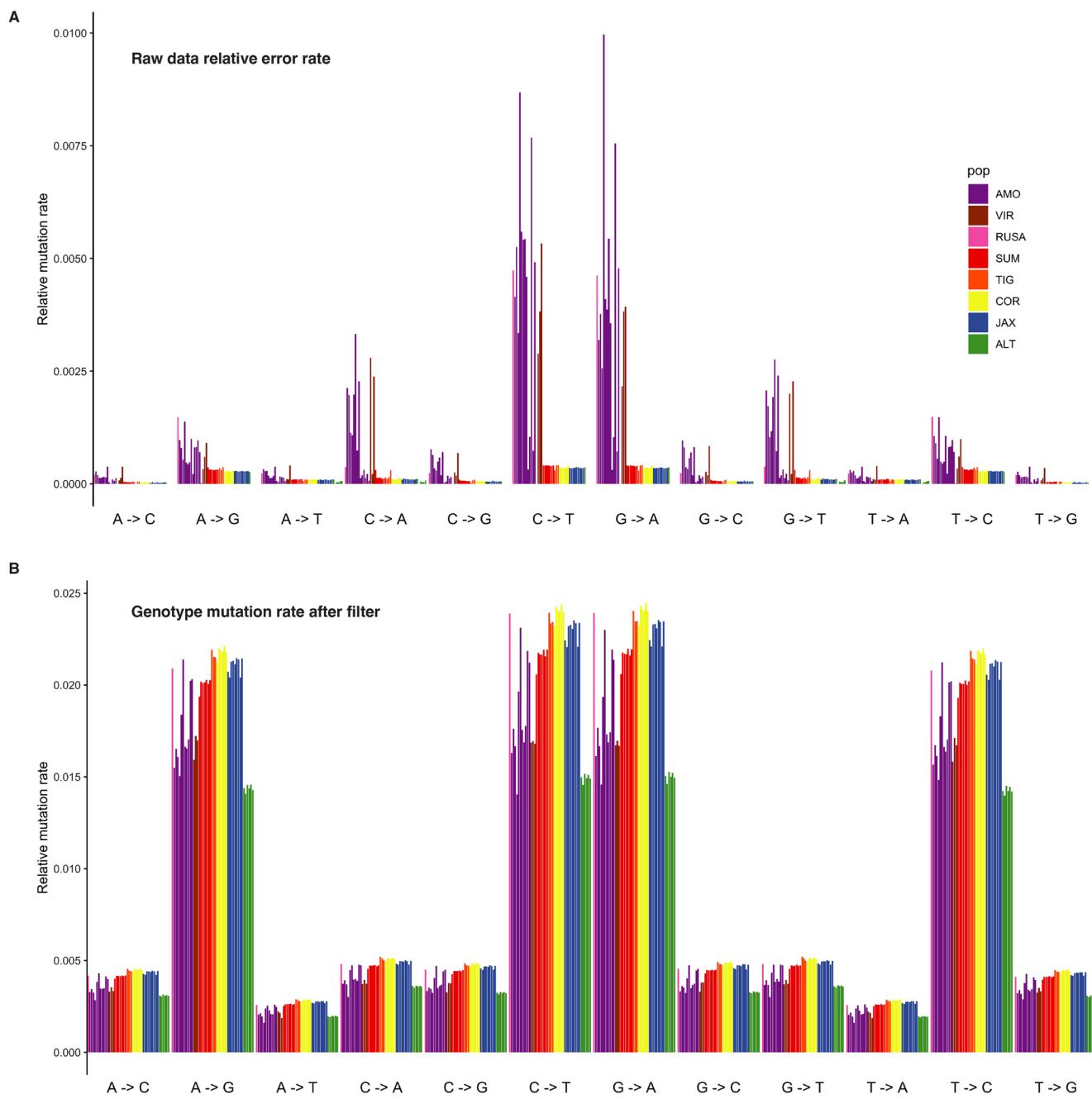
[1]The State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences; Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. [2]Federal Scientific Center of the East Asia Terrestrial Biodiversity, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, Russia. [3]Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, Yekaterinburg, Russia. [4]Ural Federal University, Yekaterinburg, Russia. [5]Section of Comparative Behavioral Genomics, National Institute on Alcohol Abuse and Alcoholism, NIH, Rockville, MD, USA. [6]Beijing Advanced Innovation Center for Genomics (ICG), Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing, China. [7]School of Archaeology and Museology, Peking University, Beijing, China. [8]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [9]Shaanxi Key Laboratory for Animal Conservation, College of Life Sciences, Northwest University, Xi'an, China. [10]Ecology and Nature Conservation Institute, Chinese Academy of Forestry, Beijing, China. [11]College of Chemistry and Molecular Engineering, Beijing National Laboratory for Molecular Sciences, Peking University, Beijing, China. [12]Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong, China. [13]University Museum, Norwegian University of Science and Technology, Trondheim, Norway. [14]Guy Harvey Oceanographic Center, Halmos College of Arts and Sciences, Nova Southeastern University, Fort Lauderdale, FL, USA. [15]Institute of Tropical Biodiversity and Sustainable Development, University of Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia. [16]Present address: Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [17]Present address: Department of Genetics, Harvard Medical School, Boston, MA, USA. [18]Present address: Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. ✉e-mail: lgdchief@gmail.com; human37564nobby@gmail.com; luo.shujin@pku.edu.cn
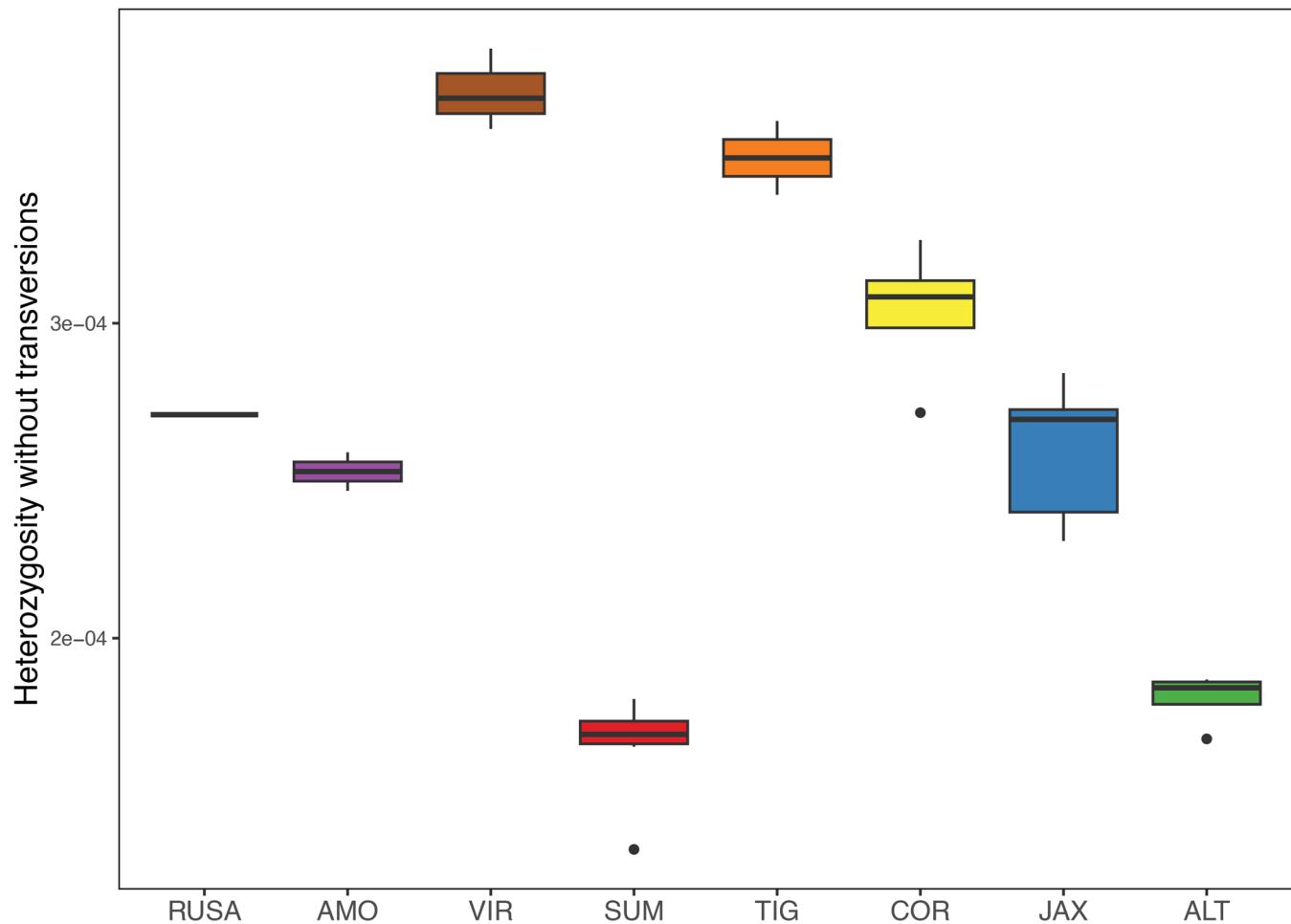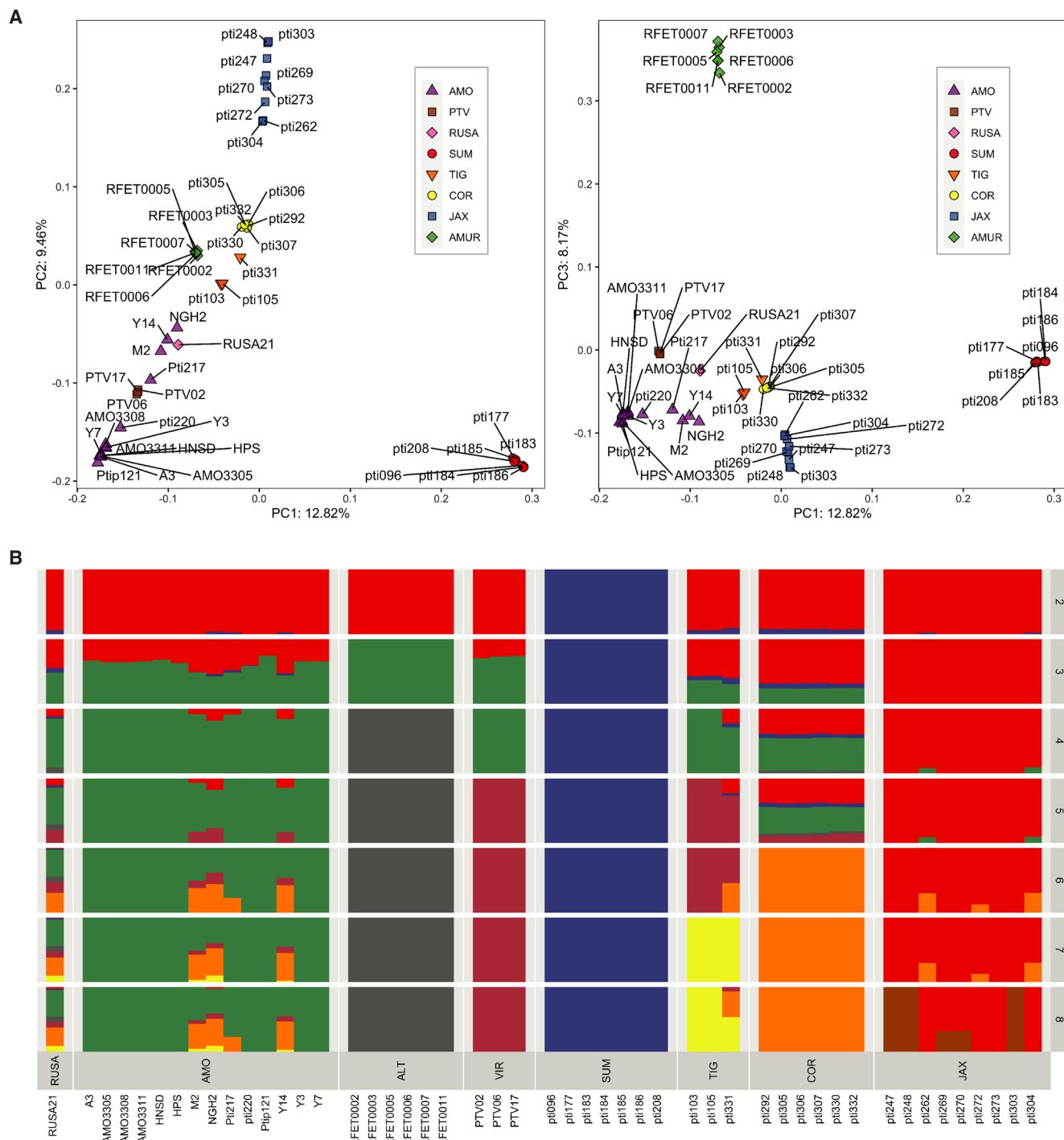
Extended Data Fig. 1 | Authentication of ancient DNA sequencing data from RUSA0021. Panels A and B show the different DNA substitution patterns at the 5′ (left) and 3′ (right) ends of reads from RUSA0021 before and after USER mix treatment, respectively. Red lines refer to C-to-T substitutions and blue lines refer to G-to-A substitutions.
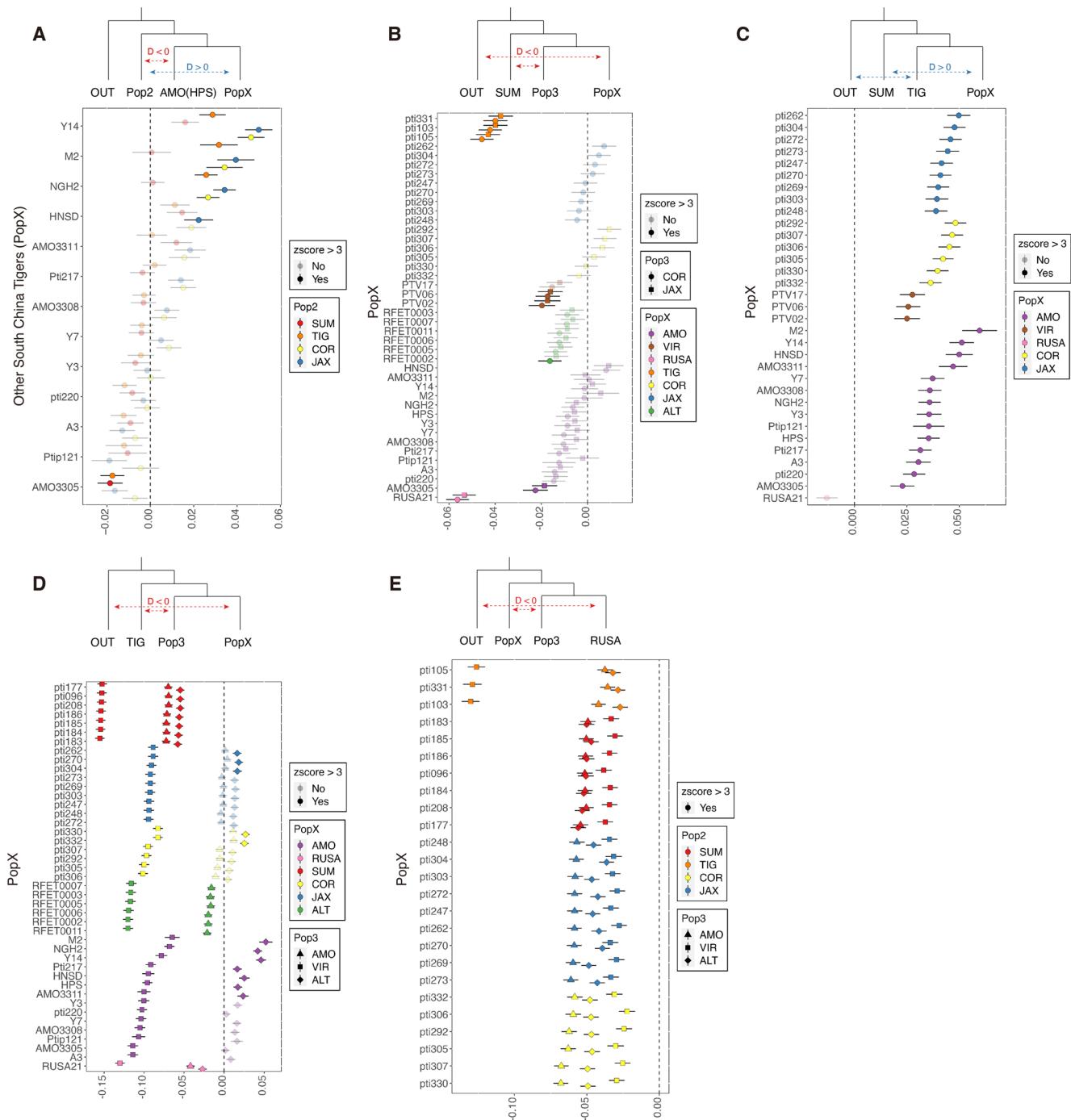
Extended Data Fig. 2 | Variable sites profiling with raw and filtered dataset. (**a**) Error rate estimation by comparing an ancient genome sequencing data to a high-quality genome data using ANGSD. (**b**) Mutation rate distribution of the filtered SNP dataset.

Extended Data Fig. 3 | Heterozygosity estimation of tigers using ANGSD. Only transitions were included for the analysis. Samples with sequencing depth lower than 4× were excluded. In the box plot, center line is the median, box bounds represent the interquartile range (IQR), whiskers extend to 1.5 × IQR from both end, and outliers are data beyond the range of whiskers.
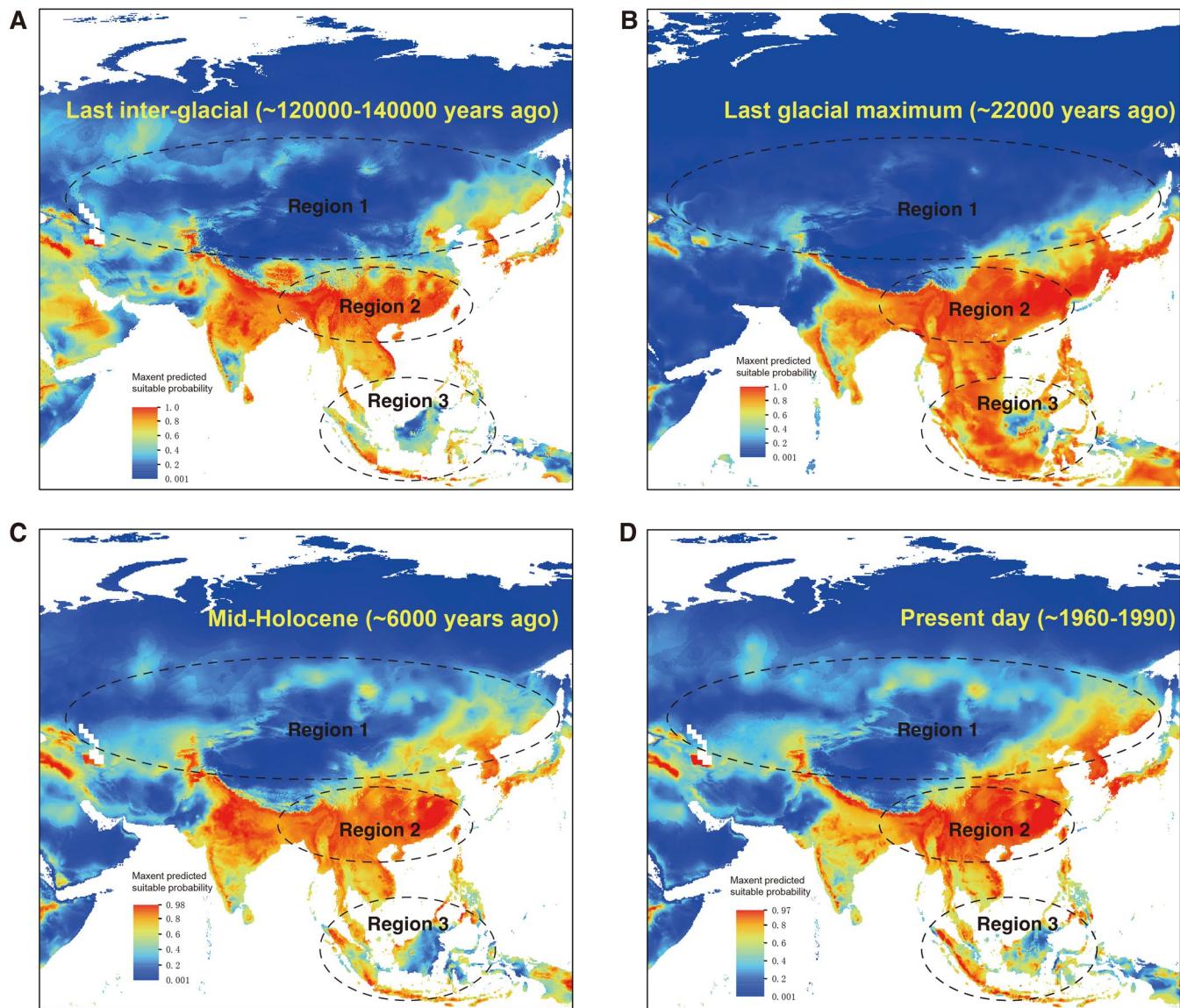
Extended Data Fig. 4 | PCA and admixture results based on genotype likelihood dataset. (**a**) The first three principal components using PCAngsd. (**b**) The supervised population structure admixture analyses using NGSadmix assuming 2 to 8 ancestral populations.
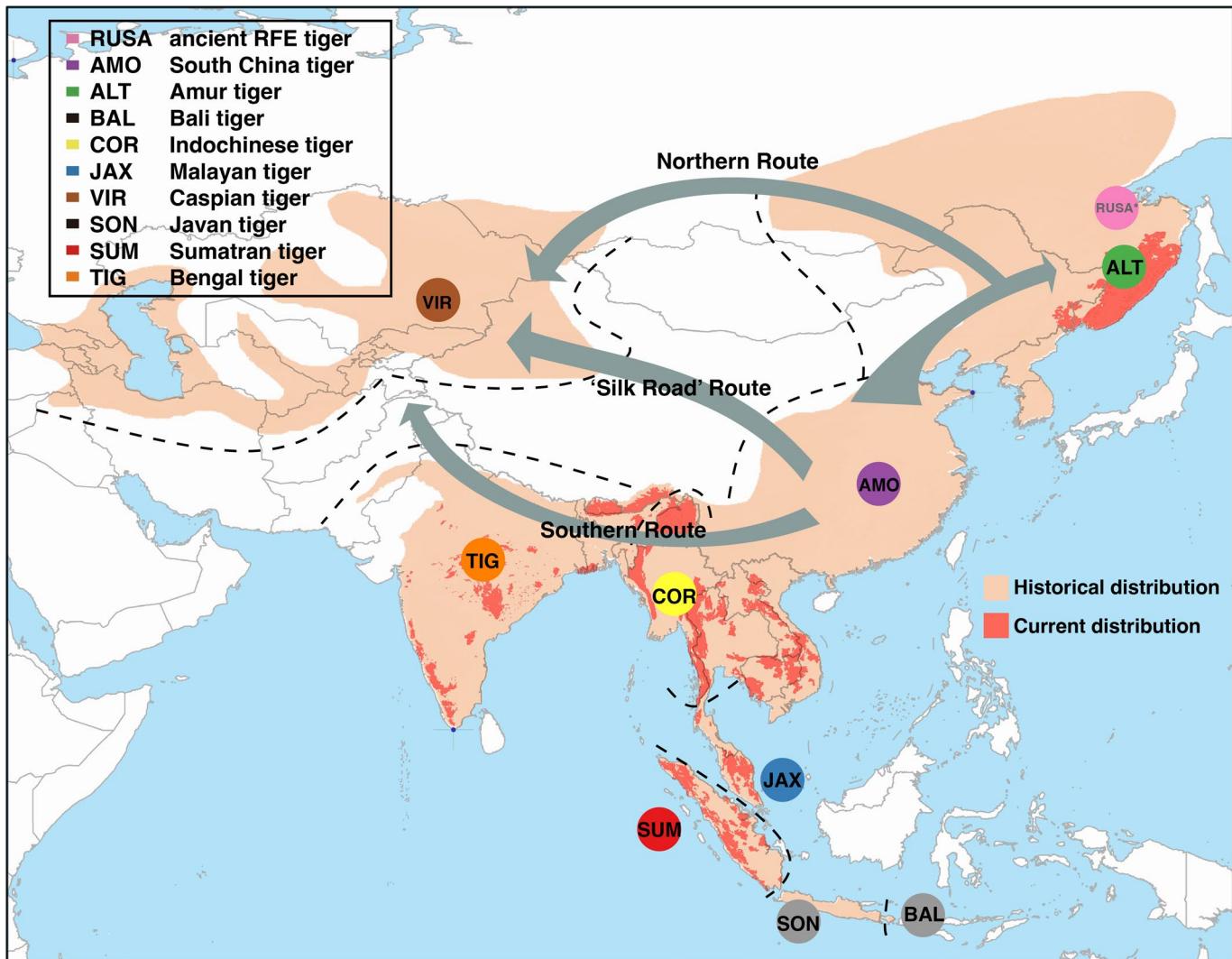
Extended Data Fig. 5 | *D*-statistic inferences for possible scenarios of gene flow between tigers. The *D*-statistic values are shown in x-axis and sample IDs of tigers representing different PopX in each testing are shown in y-axis with symbol color-coded by subspecies. Symbol shapes indicate different representatives of Pop3 in the tree topology above the *D*-statistic plot. A *D*-statistic test is considered significant with the absolute value of transformed z-score |z| > 3. Error bars show 1 × s.e.m. from a block jackknife resampling method with a block size of 5 cM. The corresponding scenarios of excessive allele sharing inferred from *D*-statistics are indicated in the tree above each plot. (**a**) Support for excessive allele sharing between different South China tiger individuals and other tiger populations (*D* > 0). (**b** and **c**) Support for excessive allele sharing between Bengal tigers and the outgroup species (*D* < 0 in B and *D* > 0 in C).

(**d**) Support for excessive allele sharing between Bengal tigers and Caspian tigers relative to that between Bengal tigers and other tiger populations (**d** < 0). (**e**) Support for excessive allele sharing between the ancient RFE tiger (RUSA) and outgroup relative to that between RUSA and other modern tigers (*D* < 0). Abbreviations for the specimens are as follows based on the geographic origin of the individual: AMO, the South China tiger (*P. t. amoyensis*); ALT, the Amur tiger (*P. t. altaica*); COR, the Indochinese tiger (*P. t. corbetti*); JAX, the Malayan tiger (*P. t. jacksoni*); SUM, the Sumatran tiger (*P. t. sumatrae*); TIG, the Bengal tiger (*P. t. tigris*); VIR, the Caspian tiger (*P. t. virgata*); RUSA, the ancient Russian Far East tiger population dated to approximately 10,000 years ago; PLE, the lion *P. leo*; PUN, the snow leopard *P. uncia*.
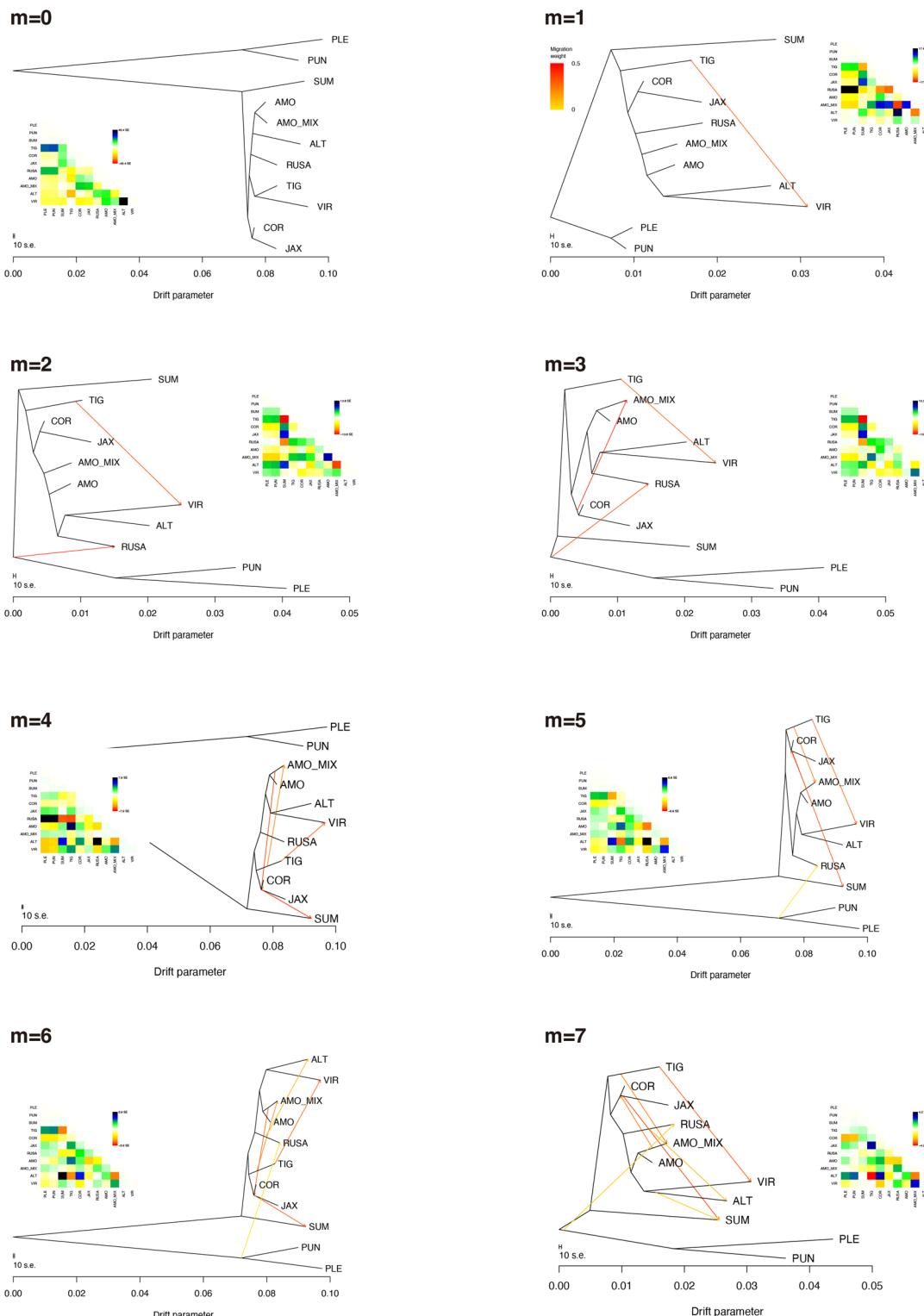
Extended Data Fig. 6 | Ecological niche modeling results of the tiger distribution during four different periods. (a) The Last Interglacial period (LIG, approximately 120,000–140,000 years ago), **(b)** the Last Glacial Maximum (LGM, approximately 22,000 years ago), **(c)** the mid-Holocene (approximately 6,000 years ago), and **(d)** present day (data from approximately 1960–1990).
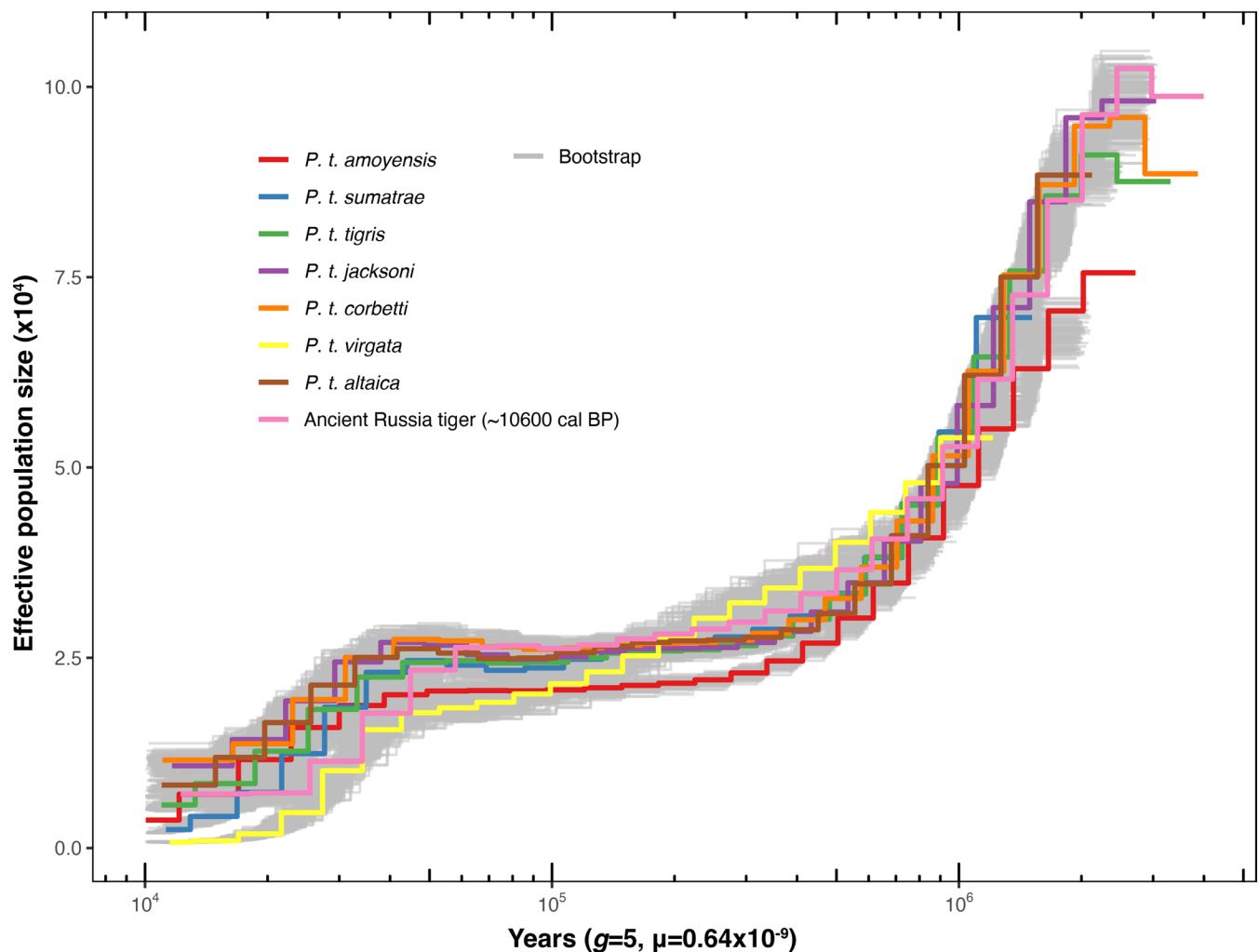
Extended Data Fig. 7 | Three postulated scenarios concerning the ancient colonization of Central Asia and the establishment of the Caspian tiger. The three dispersal routes included a southern route via the Indian subcontinent, a northern route via the Siberian plain, and a historical 'Silk Road' route through the Gansu corridor in Northwest China. Phylogenomic and demographic analyses and biogeographic modeling supported a possible initial expansion from East Asia to the modern range in Central Asia via the northern Siberian route, followed by subsequent gene flow from the ancient Bengal tiger counterpart through the Himalayan corridor.

Extended Data Fig. 8 | TreeMix phylogeny of ancient and modern tigers.
Samples were grouped by modern subspecies or ancient RFE population.
South China tigers with admixed ancestry were assigned as a separate group
(AMO_MIX). Migration edges (m) were inferred from 0 to 7. The migration
band inference results were similar to the results for gene flow and population
admixture obtained with *D*-statistics and admixture graph modeling. We
used the topology with 7 inferred migration bands for further demographic
history modeling.

Extended Data Fig. 9 | Demographic history analysis for different tiger subspecies estimated in PSMC. PSMC was applied for each tiger, and one for each subspecies is shown here. The generation time g was set to 5 years, and the mutation rate μ was calculated to be 0.64×10^{-9} substitutions per site per year. Support values from 100 bootstrap replicates for each run are shown in gray.

Extended Data Table 1 | Ancient and modern tiger samples of the study including all nine modern subspecies and ancient Russian Far East population

Scientific name	Common name	Code	Sample quality	Sample age	Sample size	Number of mitogenome retrieved	Number of WGS retrieved	Source
Panthera tigris	Ancient RFE tiger	RUSA	Zooarchaeoogical/ Ancient	1,000~10,000 years	25	7	1	This study
P. t. amoyensis	South China tiger	AMO	Historical	~100 years	31	12 mitogenomes and 16 partial mtDNA sequences	12	This study
			Fresh	Modern	1	1	1	This study
			Fresh	Modern	1	1	1	[20]
P. t. virgata	Caspian tiger	VIR	Historical	~100 years	6	6	3	This study
P. t. sondaica	Javan tiger	SON	Historical	~100 years	1	1	0	This study
P. t. balica	Bali tiger	BAL	Historical	~100 years	1	1	0	This study
P. t. sumatrae	Sumatran tiger	SUM	Fresh	Modern	7	7	7	[20]
P. t. corbetti	Indochinese tiger	COR	Fresh	Modern	6	6	6	[20]
P. t. jacksoni	Malayan tiger	JAX	Fresh	Modern	9	9	9	[20]
P. t. altaica	Amur tiger	ALT	Fresh	Modern	6	6	6	[20]
P. t. tigris	Bengal tiger	TIG	Fresh	Modern	3	3	3	[20]

Corresponding author(s): Shu-Jin Luo

Last updated by author(s): Jul 24, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We collected published dataset from NCBI SRA database.
Data analysis	We have provided detailed pipeline for our study in a public available repository at GitHub (https://github.com/xinsun1/Ancient_tiger_pop_gen).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Next Generation Sequencing raw data of the tiger samples have been deposited in the Sequence Read Archive (BioProject ID: PRJNA822019). Data processing pipeline is available at https://github.com/xinsun1/Ancient_tiger_pop_gen. Processed data file is available at https://doi.org/10.5061/dryad.73n5tb324

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	No human samples were included in this study.
Population characteristics	No human samples were included in this study.
Recruitment	No human samples were included in this study.
Ethics oversight	No human samples were included in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☐ Behavioural & social sciences   ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study retrieved genomic information from ancient and recently extinct tiger populations to get a comprehensive understanding of the evolutionary history of tigers.
Research sample	Samples were collected from archaeology excavations and museum and private collections.
Sampling strategy	Destructive sampling is performed to get genetic materials.
Data collection	Genomic DNA sequencing was performed on Libraries were sequenced on an Illumina HiSeq 2500 or X Ten platform at BIOPIC, Peking University, or Novogene Co., China.
Timing and spatial scale	Our collected sample span from ~10kya to present. Radio-carbon dating was perform to date our ancient specimens.
Data exclusions	We excluded samples with unsuccessful genomic information retrieved (endogenous DNA content lower than 1%).
Reproducibility	For our ancient samples, we prepared our DNA extraction and sequencing in different facilities, one at Peking University in China and another at University of Copenhagen in Denmark. For our museum collection samples, multiple sequencing libraries were built.
Randomization	We categorized our samples based on the population structure inferred with genetic data.
Blinding	There is no blinding for our study. Basically, our samples were labeled with its geographical origin and genetics ancestry.

Did the study involve field work?   ☐ Yes   ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging