Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection

Ellie Armstrong[1*¥], Anubhab Khan[2*], Ryan W Taylor[1,3], Alexandre Gouy[4,5], Gili Greenbaum[1], Alexandre Thiéry[4,5], Jonathan TL Kang[1,6], Sergio Redondo[1], Stefan Prost[1], Gregory Barsh[1,7], Christopher Kaelin[8], Sameer Phalke[9], Anup Chugani[9], Martin Gilbert[10,11], Dale Miquelle[10], Arun Zachariah[12], Udayan Borthakur[13], Anuradha Reddy[14], Edward Louis[15], Oliver A. Ryder[16], Yadavendradev V Jhala[17], Dmitri Petrov[1], Laurent Excoffier[4,5], Elizabeth A Hadly[1*], Uma Ramakrishnan[2*¥]

[1] Department of Biology, Stanford University, Stanford, California, USA

[2] National Centre for Biological Sciences, TIFR, Bangalore, India

[3] End2End Genomics, LLC, Davis, California, USA

[4] Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

[5] Swiss Institute of Bioinformatics, Lausanne, Switzerland

[6] Genome Institute of Singapore, A*STAR, Singapore

[7] HudsonAlpha Institute for Biotechnology, Hunstville, Alabama, USA

[8] Department of Genetics, Stanford University, Stanford, California, USA

[9] Medgenome labs limited, Bangalore, India

[10] Wildlife Conservation Society, Russia Program, New York, USA

[11] College of Veterinary Medicine, Cornell University, Cornell, USA

[12] Kerala Forest Department, Sulthan Bathery, Waynad, India

[13] Aranyak, Guwahati, India

[14] Laboratory for Conservation of Endangered Species, CCMB, Hyderabad, India

[15] Department of Genetics, Omaha Zoo, Omaha, USA

[16] San Diego Zoo, Institute for Conservation Research, Escondido, California, USA

[17] Wildlife Institute of India, Dehradun, India

*equal contribution

[¥] Corresponding author

Abstract

Tigers are among the most charismatic of endangered species, yet little is known about their evolutionary history. We sequenced 65 individual genomes representing extant tiger geographic range. We found strong genetic differentiation between putative tiger subspecies, divergence within the last 10,000 years, and demographic histories dominated by population bottlenecks. Indian tigers have substantial genetic variation and substructure stemming from population isolation and intense recent bottlenecks here. Despite high genetic diversity across India, individual tigers host longer runs of homozygosity, potentially suggesting recent inbreeding here. Amur tiger genomes revealed the strongest signals of selection and over-representation of gene ontology categories potentially involved in metabolic adaptation to cold. Novel insights highlight the antiquity of northeast Indian tigers. Our results demonstrate recent evolution, with differential isolation, selection and drift in extant tiger populations, providing insights for conservation and future survival.

Introduction

Species are classified as endangered based on recent trends in their population sizes and habitat quality (e.g. IUCN red list criteria, Mace et al., 2008). Endangerment status spurs funding, conservation action, and management in an attempt to secure species survival. Implicit assumptions underpinning risk category designations are that recent demographic trends determine extinction probability and that loss of genetic diversity and inbreeding in small populations compromises their fitness. Supporting these assumptions, empirical, theoretical, and experimental studies suggest that individual and population survival is contingent on genetic variability (Saccheri et al., 1998). Standing genetic variation in a population is determined by the interplay of mutation rate, demography, gene flow/connectivity, selection pressures over time, and genetic drift (Ellegren & Galtier, 2016). For endangered species characterized by long-term population decline, small and fragmented populations, potentially differential selection, and frequent mating between close relatives, populations could have unique histories resulting in low, but distinct standing genetic variation. If populations remain connected despite landscape fragmentation, and in the absence of differential selection, they could have shared standing genetic variation. Importantly, populations and landscapes within species distributions might have diverse histories, and hence differential probabilities of survival contingent on standing genetic variation.

Up to now, genetic diversity has been used as a proxy for evolutionary divergence, without considering whether such genetic divergence is a result of adaptation to local environments or stochastic drift, or both. Such understanding has been elusive until recently because estimating recent history of populations requires

large genomic sampling across populations for high statistical power and appropriate techniques for detection of recent selection across the genome (Pool et al., 2010). Recent advances in sequencing technology, growth of population genomic models, and enhanced computing power have revolutionized our ability to read entire genomes, allowing quantification of the sum total of genetic variation within individuals and populations.

For several endangered species, whole genome re-sequencing has revealed low species-level variation (e.g. lynx, Abascal et al., 2016), strong signatures of population decline (e.g. mountain gorillas: Xue et al., 2015) and recent inbreeding in isolated populations (wolves, Kardos et al., 2018, Robinson et al., 2019). Genomic analyses have identified mutations pre-disposing individuals to disease (Tasmanian devil, Murchison et al 2012) as well as recent protective mutations (Tasmanian devils, Epstein et al., 2016). Finally, genomics has identified signatures of population decline in extinct species (woolly mammoth: Palkopoulou et al., 2015) and strong signatures of selection prior to extinction (Passenger Pigeons, Murray et al. 2017).

Initial studies typically sequence high-coverage genomes of a few individuals, often from ex situ collections or voucher specimens, to infer levels of variation. But to better understand population genetics of endangered species, genome sequencing efforts should be at larger scales and sample geographic landscapes comprehensively. Broader sampling is made particularly challenging in wide-ranging endangered species, especially those with geographic ranges spanning many international borders, where both sampling permissions and population management strategies differ.

The tiger (*Panthera tigris*) is an iconic and charismatic endangered species that

once spanned 70 degrees of latitude. Between 2,154 and 3,159 tigers remain, now occupying less than 6% of their 1900 A.D. range (Goodrich et al., 2015). Despite this recent range collapse, tigers still live across 11 Asian nations, and habitats that include, for example, estuarian mangrove forests (the Sundarbans), dry deciduous forests (parts of India), tropical rainforests (Malay Peninsula) and cold, temperate forests (Russian Far East). Tigers were classified into four extant (and four extinct) subspecies (Nowell and Jackson, 1996), while genetic and other data substantiated (e.g. Luo et al., 2004, suggested an additional population group) or contradicted (Wilting et al., 2015, suggested fewer population groups) this classification. Liu et al. (2018) recently presented the first analyses of genome-wide variation using voucher specimens across tiger range, and inferred relatively old divergences (~68,000 years ago) between subspecies, and low subsequent gene flow (1-10%). However, their sampling of the most populous and genetically diverse Bengal tigers was limited (in terms of habitats and numbers of samples).

60-70% of the world's extant wild tigers reside in the Indian subcontinent (Jhala et al., 2014). Limited genetic data across the tiger's range but including multiple landscapes in India, suggested high genetic diversity and differentiation within India (Mondol et al., 2009). Genome-wide studies have revealed multiple, distinct populations within India (Natesh et al., 2017). Therefore, a comprehensive understanding of tiger demographic history must include genomes sampled from the various landscapes across the Indian subcontinent. Here, we used whole genomes from across wild tiger range, with representation from all extant subspecies (except *Panthera tigris corbettii*) and most habitats to investigate (a) population clustering within range-wide samples, (b)

demographic history and differential selection in tiger populations with a subset of high-coverage samples and (c) possible signatures of recent inbreeding.

Results

We sequenced genomes from 65 individuals (Figure 1, Supplementary Table 1) at varying coverage (4.2X-32.9X, median 14.4X). Our samples included wild-caught and captive-bred tigers from four putative extant subspecies (Bengal, Malayan, Amur and Sumatran, see Supplementary Table 1 for details). We were unable to sample the South-China tiger *(P. t. amoyensis*), considered extinct-in-the-wild. While the South China tiger is thought to be ancestral, Liu et al., 2018 suggested uncertainty about the antiquity of this population, since mitochondrial genomes were similar to Amur tigers.

In order to better understand genome-wide variation and call variants reliably, we first improved the tiger genome assembly using the 10X Genomics Chromium Platform (Mohr et al. 2017) for a wild-caught Malayan individual. Based on Assemblathon2 statistics (Bradnam et al. 2013), this improved assembly corresponded to a 3.5-fold increase in the contig N50 value to 1.8 Mb and a 2.5-fold increase in the scaffold N50 value to 21.3 Mb (as compared to Cho et al. 2013; Supplementary Table 2). In addition, the resulting assembly had ~1% fewer ambiguous bases across all scaffolds (Supplementary Table 2). Details of samples used for various analyses are in Supplementary Table 3.

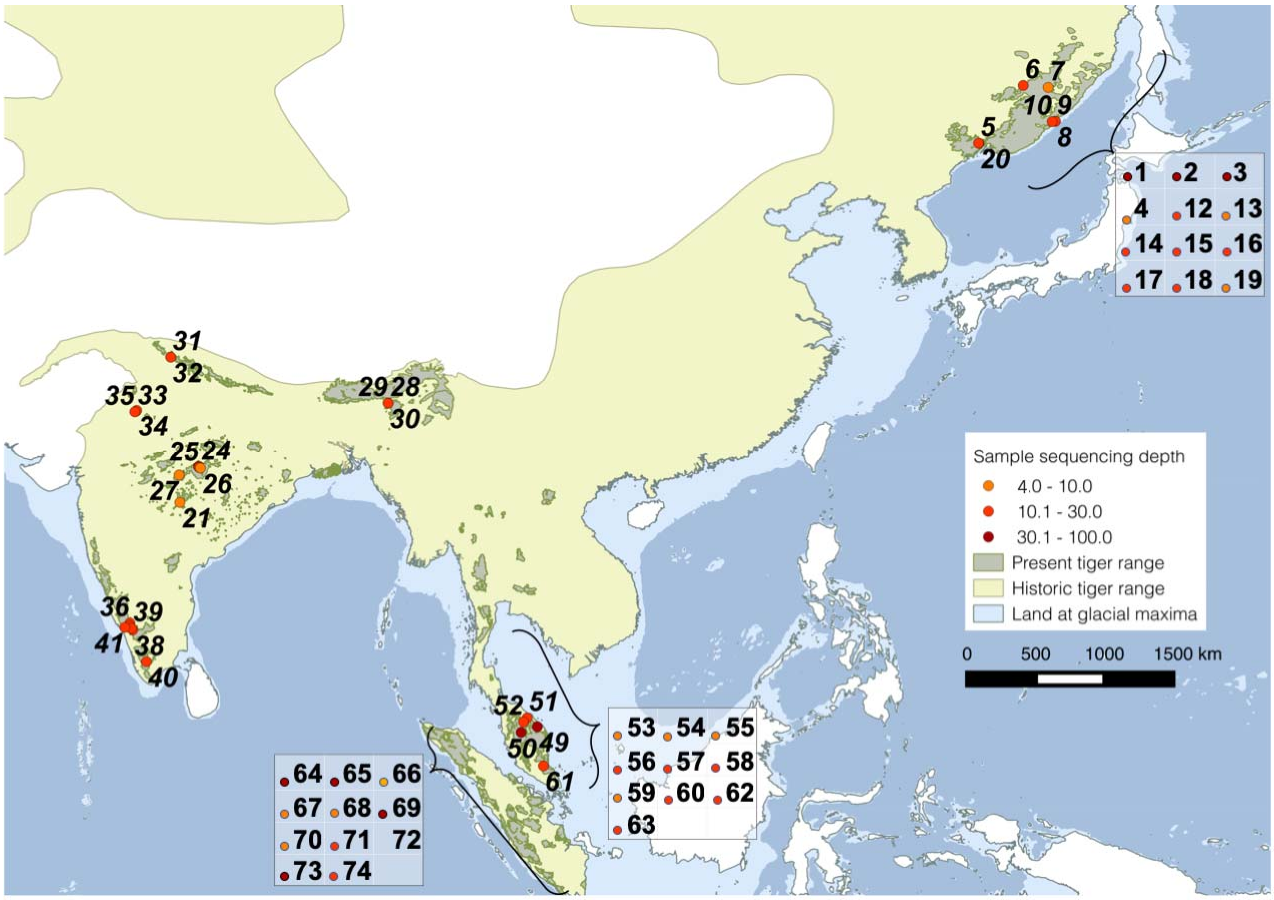

Figure 1: Tiger samples used in this study. Coverage for samples is represented by marker color. Finally, wild samples (n=32) are represented on the map, while captive individuals (n=34) are indicated in boxes by brackets. Each number refers to an individual. Sample details presented in Supplementary table 1. Historical and present range map courtesy IUCN (Goodrich et al 2015).

Population structure and genetic variation

We investigated the partitioning of range-wide tiger genomic variation using several approaches: (1) model-based, using ADMIXTURE (Alexander et al., 2009) (Figure 2a); (2) visualization with Principal Component Analyses (PCA; Chang et al. 2015) (Figure 2b); (3) F_{ST} statistics (table 1); (4) network-based analyses (Figure 2c).

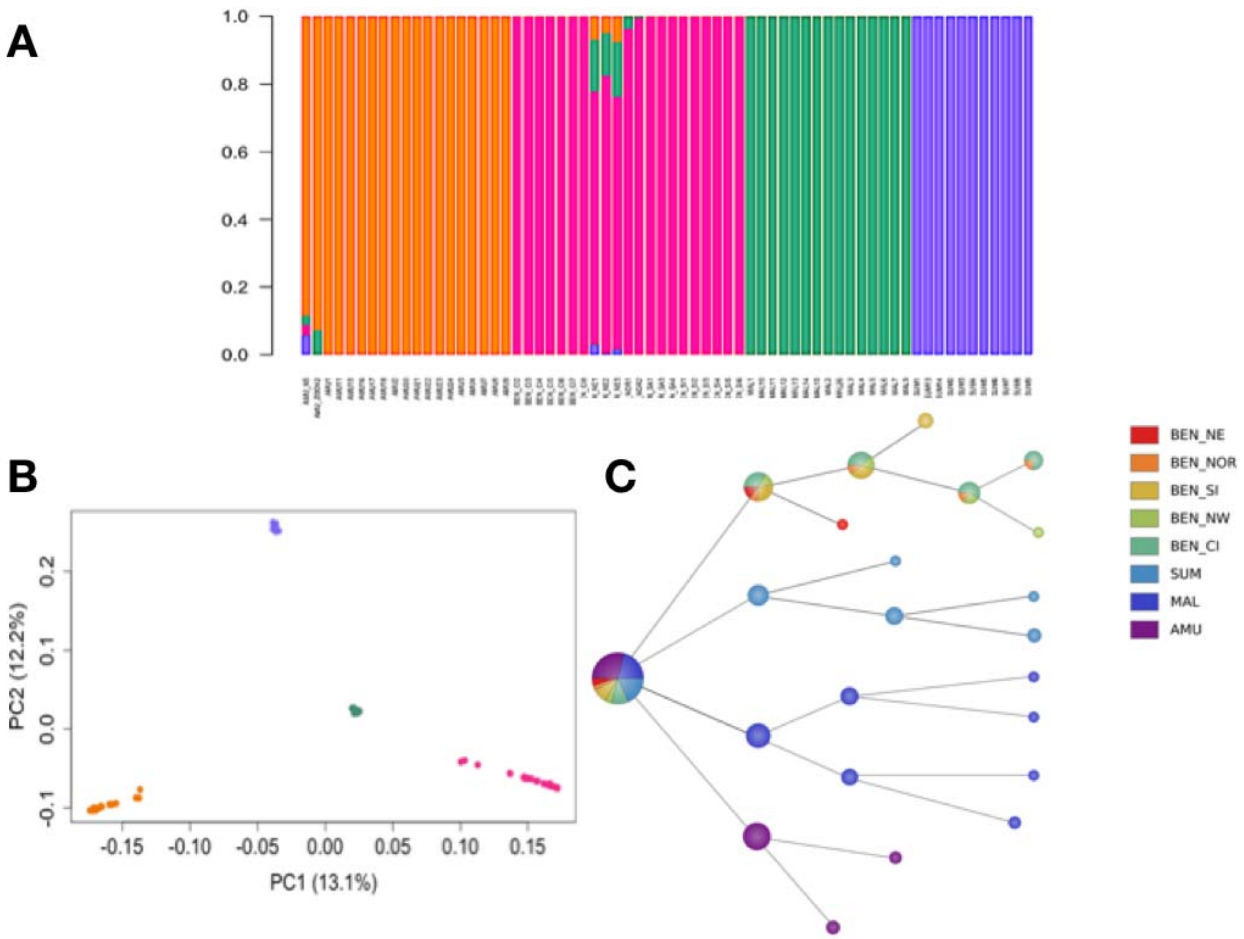

Figure 2: (a) ADMIXTURE and (b) Principal component analyses (PCA) revealing genetic population structure in tigers; and (c) branching pattern between individuals as determined by Netstruct_Heirarchy. Colors in PCA denote individuals from each population as identified by clustering in ADMIXTURE.

Model-based ADMIXTURE analyses suggested that genetically distinct populations are concordant with earlier definitions of subspecies (as also suggested by Luo et al., 2019 and Liu et al., 2018) (Figure 2A). Cross-validation statistics suggested that K=4 fit the data best (Supplementary Figure 3). At K=4, Bengal individuals sampled from the northeastern region of India show some admixture with Malayan individuals and to a lesser extent with Amur individuals. At higher K (K=5, Supplementary Figures 2 and 3) the data reveal substructure within India separating south Indian tigers from others in India, but no further substructure in the other subspecies. Higher values of K fit poorly.

PCA (Figure 2b) revealed a similar pattern to the ADMIXTURE analyses, with the subspecies separating out as clusters. PC1 separates the same four groups, and both PC1 and PC2 revealed high subdivision (PC1: 13.1%; PC2:12.2%). We henceforth refer to the geographic regions by their sub-specific names (Northeast Asia: Amur; South Asia: Bengal; Malay Peninsula: Malayan; and Sumatra: Sumatran). Additionally, PC1 shows stronger similarity between Bengal and Malayan tigers than Bengal and Sumatran tigers, consistent with the majority of results from K=3. PC2 resolved individuals in the east-to-west direction and PC1 resolves in the north-to-south direction. PC2 (12.2% variation) and PC3 (10.9% variation) further separate the four groups, and also separated of some individuals within populations (Supplementary Figures 4 and 5). Contrastingly, a PCA analysis of non-transcribed regions including only high-coverage individuals within the dataset (Sumatran=3; Bengal=3, Malayan=3, Amur=3, Supplementary Table 2) suggested that (Supplementary Figure 10) the Amur population is much less differentiated and closer to the Malayan population. Our results indicate that both Amur and Malayan populations were genetically closer to a putative ancestral Asian tiger population.

PCA within subspecies (Supplementary Figures 4 and 5) suggested that Bengal tigers clustered into four sub-groups: (1) south India, (2) central and north India, (3) north eastern India, and (4) north western India. Some genomic sub-structuring was apparent in Malayan tigers, somewhat reflective of whether tigers were sampled from the northern or southern Malayan peninsula (Supplementary Figures 4 and 5). We did not find strong signatures of population sub-structuring within Amur tigers, but we obtained only one sample from one of the sub-populations identified by Henry et al.

(2009) and Sorokin et al. (2015), Supplementary Figures 4 and 5). Within subspecies structure was confirmed in the additional PC axes for the full dataset (Supplementary Figure 6). While PC1, PC2 and PC3 separated putative subspecies (Amur, Bengal, Sumatran, and Malayan), additional axes (PC4 and higher) separated the Bengal populations according to their geographic location (north western India, south India and central, north, and north east Indian tigers comprise three distinct groups). Minimal separation occurred within Malayan populations on PC4.

Using Vctfools (Danecek et al., 2011), we estimated pairwise F_{ST}'s (Table 1) between each of the four subspecies.

Table 1: Weighted F_{ST} between subspecies as computed by VCFtools.

Population	Bengal	Malayan	Sumatran
Amur	0.200	0.230	0.318
Bengal		0.164	0.242
Malayan			0.280

F_{ST} values were approximately equal between subspecies, differences were consistent with geography. The F_{ST} between the Malayan and Bengal groups (0.164) was lowest, while Amur and Sumatran F_{ST} (0.318) was highest, consistent with both the ADMIXTURE and PCA. F_{ST} between putative Bengal tiger subpopulations in India (Supplementary Table 5) revealed high subdivision.

Branching patterns in a population structure tree (Figure 2c) generated by Netstruct_Hierarchy (Greenbaum et al., 2019), suggested that differentiation in tigers corresponded to the four putative subspecies. The analysis also reflected substructure within Bengal tigers, with the northeastern population being the most distinct, followed

by subdivision of south Indian tigers, a central and northern Indian tiger group, and a north-western Indian tiger group, similar to the PCA results described above.

We compared genome-wide variability between tiger subspecies/subpopulations to other cats (N=7) and endangered species (N=8, including endangered cats). Tigers had relatively high species-level genetic diversity (Supplementary Figure 7). However, different tiger subspecies (Amur, Bengal, Malayan, and Sumatran), or individuals from the same subspecies had different levels of SNV diversity. In other words, some individuals were more homozygous on average than others, even within a subspecies, suggesting that a voucher specimen-based approach to estimate extant diversity might provide an incomplete picture.

Bengal tigers had the highest nucleotide diversity (Supplementary Table 6), while Sumatran tigers had the lowest. Because Bengal tigers had disproportionately high sample sizes, we conducted a rarefaction analysis (ADZE; Szpiech et al. 2008) which revealed that diversity estimates were approaching saturation for all populations (Supplementary Figure 8). Rarefaction re-iterated that Bengal tigers had the highest variation for both private or unique genetic variation.

Demographic history of subspecies

We first reconstructed the past demographic history of each population with PSMC (Pairwise sequentially Markovian coalescent; Supplementary Figure 9), and our results paralleled those in Liu et al. 2018: all populations of tigers exhibit similar evolutionary patterns of population size decline.

Simulated site frequency spectrum (SFS) based on coalescent modeling allowed us to investigate subspecies divergence, population size changes, as well as gene flow. The best fit scenarios supported a very recent (Holocene) divergence of all tiger subspecies (Figure 3) from an ancestral population. Simulations supported a very strong bottleneck occurring around 234,000 years ago, with most remaining lineages coalescing rapidly, consistent with a speciation event. This timing was consistent with signatures of population decline in the PSMC analysis (Supplementary Figure 9).

Existence of a large (theoretical) Asian metapopulation of tigers was followed by recent divergences between all four subspecies and between populations within the subspecies, including those within India. The best-fit scenario supported subspecies divergence in the Holocene between 7,500 and 9,200 years ago (i.e. 1,500 and 1,840 tiger generations). Sumatran tiger divergence correlates with sea levels rise (Heaney, 1991) and separation of the island of Sumatra (but we imposed that this divergence post-date the last-glacial maximum, i.e. 18,000 years ago or younger). Estimated migration rates were very low, with all populations receiving fewer than one migrant per generation; populations have been quite isolated since their initial early Holocene divergences. Additionally, we found that Sumatran and Bengal populations show evidence for signals of a founding bottleneck, but Amur and Malayan populations do not. Both Sumatran and Amur tigers also showed evidence of strong recent bottlenecks.

We further modeled the divergence within Bengal tigers into four populations: Northwestern India, central India, southern India, and northeastern India. Because PCA and Netstruct suggest that central and north Bengals are a single population (and north Indian tigers were not sequenced at high enough coverage), we did not include them in

demographic analyses. We assessed the robustness of the northeastern population being a part of the Bengal subspecies. In order to do so, the northeast population was modelled as an independent subspecies, and allowed to diverge directly from the Asian metapopulation. However, such a model is a poorer fit to the data than if northeast Indian tigers are considered part of the Bengal subspecies ($\log_{10}$Likelihood difference between model is 37). Within Bengal tigers, divergences are extremely recent (within the last 2,000 years), except for the northeast, which diverged early (6,800 years ago) after the separation of Bengal tigers 8,400 years ago from the ancestral Asian metapopulation. Within India, the northwestern population underwent a strong bottleneck at the time of its founding. Recent bottlenecks were most severe in the northwest and southern populations, while the northeastern and central populations showed relatively weaker bottlenecks. Overall, tiger populations from all subspecies revealed signals of strong recent bottlenecks except central and northeastern Bengal tigers.

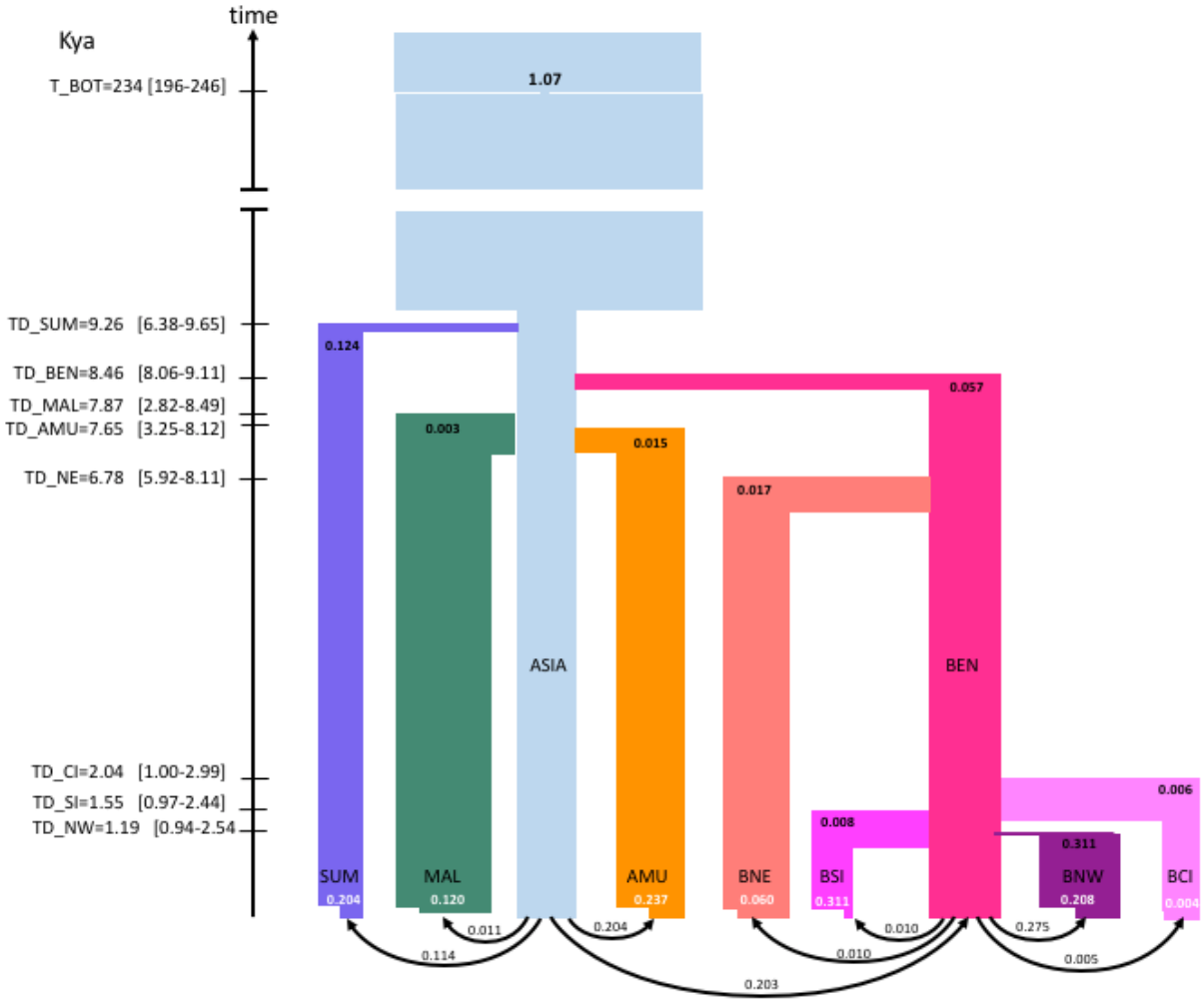

Figure 3: **Estimated demographic history of Asian tigers**
Sumatra (SUM: lavender), Malayan (MAL: dark green), Amur (AMU: orange) and ancestral Bengal (BEN: hot pink). Subspecies are assumed to have diverged from an ancestral Asian (Asia, light blue) metapopulation, sometime in the past, while allowing continuous gene flow since their separation (curved arrows and text in units of 2Nm). The Bengal tigers further differentiated into North East (BNE, salmon pink), Central (BCI, light pink), Southern (BSI, dark pink) and North West (BNW, purple) populations (approximate geographical locations of landscapes, modified from Natesh et al., 2017, shown in the inset map), still receiving some gene flow from the ancestral (theoretical) Bengal metapopulation. Founder effects at separation times were modelled as instantaneous bottlenecks with intensities (t/2N, reported in black and white text respectively), and represented as horizontal lines with widths inversely proportional to intensity. Recent population contractions were also implemented as instantaneous bottlenecks with intensities (t/2N, reported in white text) inversely proportional to current population size. Population widths are approximately proportional to estimated population sizes. The ancestral (theoretical) Asian tiger population was assumed to have a larger population size than the current (theoretical) Asian metapopulation, and to have gone through an instantaneous bottleneck sometime in the past. Divergence (T_DIV) and bottleneck times (T_BOT) are reported in ky (thousand years ago),

assuming a mutation rate of $0.35*10^{-8}$ and 5 years per generation. Times 95% CI values are shown within brackets on the left of the time arrow. Estimated values and associated 95% CI of all parameters are reported in Supplementary Tables 7 and 8.

Genome scans for selection

We investigated how genetic patterns might have been impacted by natural selection in the four tiger subspecies (Amur, Bengal, Malayan, and Sumatran). We computed a statistic, mPBS (metapopulation branch statistic, a simple extension of the PBS statistic of Yi et al. (2010), see Material and Methods), measuring the length of the branch leading to a given subspecies since its divergence from the others (Fig 4a and Material and Methods). mPBS is similar to other traditional measures of genetic differentiation (e.g. F_{ST}) used in genome scans: higher than average values signify differential positive selection. Extreme mPBS values should thus correspond to regions that have been targeted by natural selection.

The genome-wide distributions of the mPBS revealed that Bengal and Malayan populations had the lowest average values, suggesting short terminal branches subsequent to the divergence of these two populations from the hypothetical metapopulation (Fig 4b-f). On the contrary, Amur and Sumatran tigers had high values on average (Fig 4b-f).

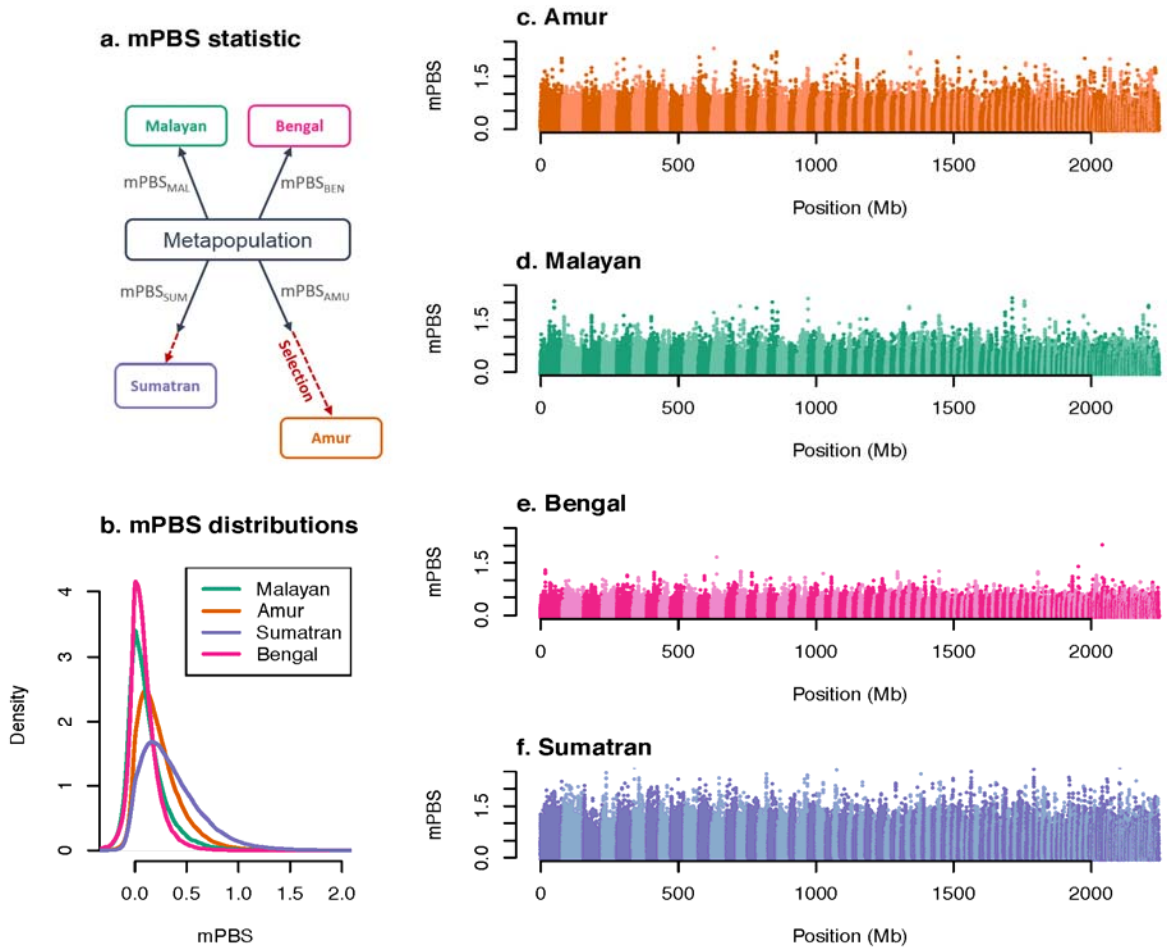

FIGURE 4: Genome scan for selection. (a) We present the mPBS statistic with a hypothetical model where the 4 populations diverge from a metapopulation, and where selection acts in both the Amur and Sumatra lineages, (b) the global distribution of observed mPBS for each population. Panels c to f correspond to the genome-wide distributions of the statistic for (c) Amur, (d) Bengal, (e) Malayan and (f) Sumatran tigers as a function of the genomic position. Alternating light and dark colors indicate different scaffolds.

We observed little difference between transcribed and non-transcribed regions in mPBS distributions, suggesting no strong differential impact of background or positive selection in tiger coding regions (Supplementary Figure 11). Both tails of the distribution were enriched (we did not filter for mutation types), possibly caused by biased gene conversion (Supplementary Figure 11). Average mPBS values were higher when

considering only individuals with average coverage > 10X than when comparing fewer individuals with highest coverage (Supplementary Figure 11).

Overall, the mPBS distribution obtained under the neutral demographic model (Figure 4) fit very well with the observed distribution (Supplementary Figure 11), implying that most observed differences between populations could be explained by their demographic history. We predicted high mPBS values in Amur tigers and Sumatran tigers where small effective sizes would yield high levels of genetic drift, but the observed values are even higher than those expected (Supplementary Figure 12), suggesting a possible effect of natural selection on genomic diversity in these subspecies. In contrast, we observed no apparent deviation of observed mPBS values from a purely neutral model in Bengal and Malayan populations.

Enrichments tests were then used to detect targets of selection. These tests are a conservative approach to detect selection because they are less susceptible to the influence of non-selective forces. The observation of an excess of moderately high values in Amur and Sumatran tigers rather than a few very extreme values is compatible with the effect of polygenic selection rather than hard selective sweeps. We looked for signal of polygenic selection using functional enrichment tests (Daub et al. 2013, Gouy et al. 2017) in an attempt to identify biological functions putatively targeted by selection. We used mPBS values computed on all individuals (Amur and Sumatran) with average coverage greater than 10X and mapped the top 0.1 % regions with highest mPBS values to annotated genes (+/- 50 kb flanking regions). 119 and 80 genes (in Amur and Sumatran tigers, respectively) were found within these top 0.1% regions. We identified 15 statistically significant Gene Ontology (GO) terms in Amur tigers, and 5 in

Sumatran tigers (Table 1). Out of the 15 GO categories identified, 4 have an unspecific function and the 11 others are involved in lipid processing and metabolism (Table 2).

The genes responsible for the enrichment in fat metabolism-related GO terms were all included in the Cellular lipid metabolic process (GO:0044255). These included, for example, the Apolipoprotein B receptor (APOBR) or Caveolin-1 (CAV1) that are involved in the modulation of lipolysis. Fat metabolism enzymes included Phosphatidate phosphatase (LPIN2), Phospholipase B-like 1 (PLBD1), and Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 2 (HACD2). We also identified genes involved in the mitochondrial respiratory chain: a Cytochrome P450 subunit (CYP1A2) and the mitochondrial Lipoyl synthase (LIAS). Cardiolipin synthase (CRLS1) is involved in the synthesis of cardiolipin, an important phospholipid of the mitochondrial membrane critical to mitochondrial function. Finally, Thromboxane-A synthase (TBXAS1) is involved in vasoconstriction and blood pressure regulation.

In Sumatran tigers, significant GO terms were related to cell development regulation: Regulation of neuron projection development (GO:0010975), Regulation of anatomical structure size (GO:0090066), and Regulation of cell development (GO:0060284). These four terms contain the same 6 genes: Tyrosine-protein kinase (RYK), E3 ubiquitin-protein ligase (RNF6), Prolow-density lipoprotein receptor-related protein 1 (LRP1), Angiotensin-converting enzyme (ACE), Rap1 GTPase-activating protein 2 (RAP1GAP2) and B2 bradykinin receptor (BDKRB2). These genes are involved in morphological development, and selection targeting these loci could be good candidates for size differences between Sumatran tigers and their relatives. Two other

terms are significant, related to toxic substance processing: Response to toxic substance (GO:0009636) and Organophosphate biosynthetic process (GO:0090407).

TABLE 2: Gene Ontology enrichment results. The 20 most significant GO terms are presented, as well as their total number of genes, the number of observed significant genes in a given term, the expected number of significant genes, the fold enrichment and the p-value of the Fisher's exact test.

Population	GO term	Number of genes	Observed	Expected	Fold enrichment	P-value
Amur	Cellular lipid metabolic process (GO:0044255)	675	15	6.61	2.27	0.0037
Amur	Sulfur compound biosynthetic process (GO:0044272)	120	5	1.17	4.26	0.0073
Amur	Carboxylic acid biosynthetic process (GO:0046394)	225	7	2.2	3.18	0.0074
Amur	Organic acid biosynthetic process (GO:0016053)	226	7	2.21	3.16	0.0075
Amur	Lipid metabolic process (GO:0006629)	822	16	8.05	1.99	0.0090
Amur	Lipid biosynthetic process (GO:0008610)	424	10	4.15	2.41	0.0095
Amur	Glycerolipid metabolic process (GO:0046486)	299	8	2.93	2.73	0.0100
Amur	Membrane organization (GO:0061024)	496	11	4.86	2.27	0.0102
Amur	Phospholipid metabolic process (GO:0006644)	305	8	2.99	2.68	0.0111
Amur	Vesicle organization (GO:0016050)	192	6	1.88	3.19	0.0126
Amur	Phospholipid biosynthetic process (GO:0008654)	197	6	1.93	3.11	0.0141
Amur	Monocarboxylic acid biosynthetic process (GO:0072330)	145	5	1.42	3.52	0.0153
Amur	Receptor-mediated endocytosis (GO:0006898)	146	5	1.43	3.5	0.0157
Amur	Lipid localization (GO:0010876)	212	6	2.08	2.89	0.0194
Amur	Glycerophospholipid biosynthetic process (GO:0046474)	164	5	1.61	3.11	0.0243
Sumatran	Regulation of neuron projection development (GO:0010975)	313	6	2.06	2.91	0.0179
Sumatran	Response to toxic substance (GO:0009636)	320	6	2.11	2.85	0.0197
Sumatran	Organophosphate biosynthetic process (GO:0090407)	414	7	2.73	2.57	0.0200
Sumatran	Regulation of anatomical structure size (GO:0090066)	325	6	2.14	2.8	0.0211
Sumatran	Regulation of cell development (GO:0060284)	621	9	4.09	2.2	0.0214

Runs of Homozygosity

Historical demography and recent inbreeding are detectable through runs of homozygosity (ROH) in the genome (Kirin et al 2010, Pemberton et al 2012, Kardos et al 2018). We quantified long (>1Mb) and short (10-100kb, 100kb-1MB) homozygous stretches, as well as the proportion of ROH in the genome for several individuals

(Figure 5). All individuals showed high frequency of short stretches (Figures 5A & B), possibly due to common recent bottlenecks (Ceballos et al., 2018). Somewhat surprisingly, individuals from the demographically large Indian tiger population revealed a high proportion of their genomes in long ROH. A closer look revealed that the inbred Bengal tigers are predominantly from very small isolated sub-populations. Tigers from the large central India population had lower values of ROH, while some south Indian tigers (Periyar Tiger Reserve; BEN_SI5) and north-west Indian tigers (Ranthambore Tiger Reserve; BEN_NW3, BEN_NW4) potentially had the highest levels of recent inbreeding (ROH > 1MB. We verified these results using a sliding window approach (Supplementary Figure 13).

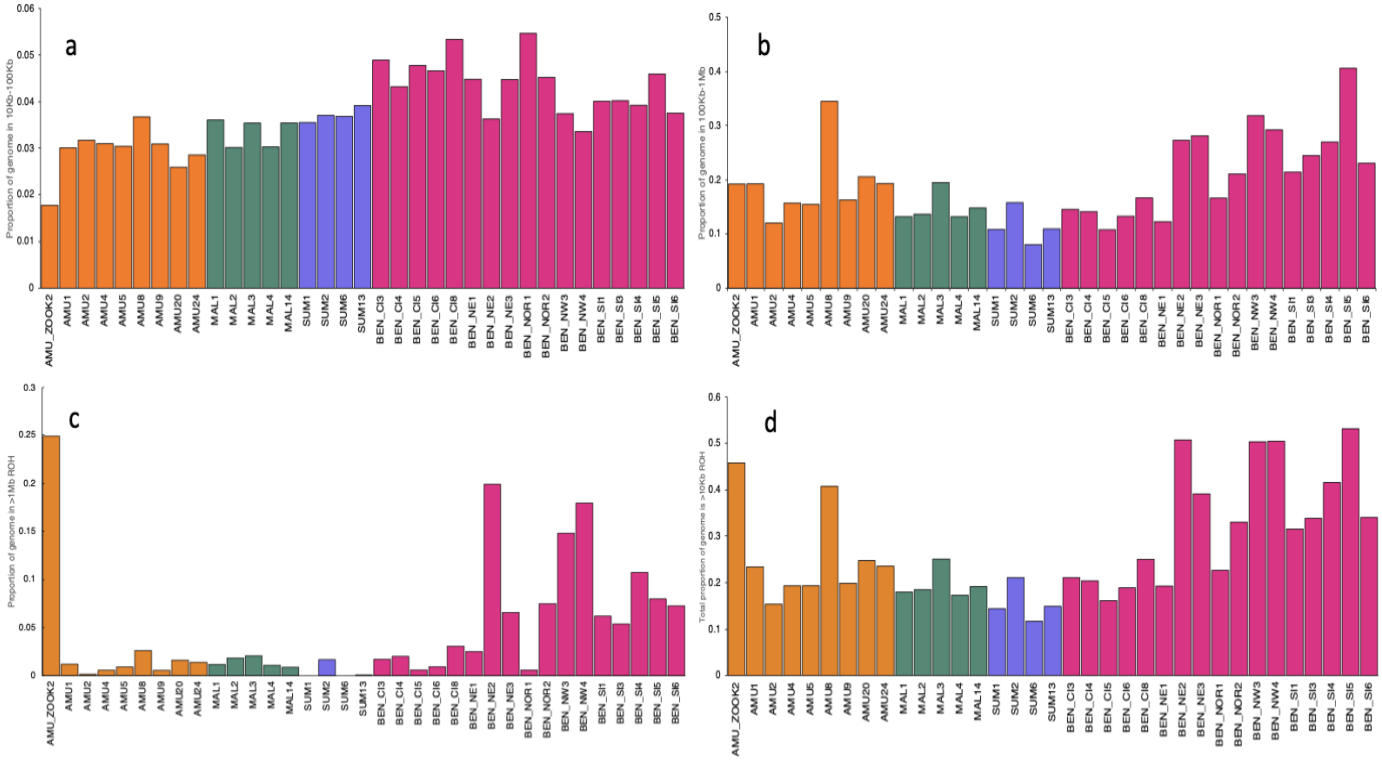

Figure 5: Runs of homozygosity inferred based on different run lengths, (a) 10-100kb, (b) 100kb-1Mb, (c) above 1 MB and (d) Total ROH, which includes all run lengths greater than 10kb.

Discussion

What do genomes tell us about the history of the tiger?

We sampled genomes from four of five extant wild tiger subspecies, augmented by extensive sampling within the Indian subcontinent, which contains most of the world's remaining wild tigers. We were unable to sample Indochinese tigers *(P. t. corbetti).* Because tigers have such a large but continuous geographic range, we expect signatures of population structuring (Luo et al., 2004, Luo et al., 2019, but see Wilting et al., 2015). The variety of analyses we conducted (model-based inference, PCA, network-based inference, F_{ST}, demographic modeling) revealed that tigers from different geographical locations are genetically distinct and have been predominantly isolated from each other for 8,500 to less than 2,000 years (Figure 3). These patterns may reflect loss of connectivity due to sea level rise, which has separated the formerly continuous Sahuel subcontinent of southeastern Asia into isolated islands, and changing environments due to human population size increase, the rise of agriculture, and climatic change of the mid-late Holocene. Although the timing and severity of the events differentiating tiger subspecies differ, our data and analyses confirm previous inferences (Liu et al. 2018) that the four tiger putative subspecies are valid entities both geographically and genetically, and that post-divergence gene flow has been relatively low. Theoretical predictions (based on body size, Sutherland et al., 2000) and empirical results (genetics: Joshi et al., 2013, Yumnam et al., 2014; camera trap: Singh et al., 2013) suggest that tigers can move extraordinary distances (e.g. 300 km), even across human-dominated landscapes. Such long-range movement might result in relatively low genetic differentiation. Re-iterating, despite the possibility of long-distance dispersal, our

models suggest that migration rates between tiger populations have been relatively low recently, emphasizing separate recent evolutionary histories. Overall, our analyses (and Liu et al., 2018) contradict currently accepted IUCN management criteria (Kitchner et al., 2017).

Because we sampled across landscapes within subspecies, we were able to compare population structuring within the four subspecies. Population structure within tiger subspecies has been illustrated before (e.g. Sorokin et al., 2016; Natesh et al. 2017; Thapa et al., 2018), and our results conclusively reveal population structure between subspecies, but also significant structure within some subspecies. We showed that population genetic substructure is highest in the Indian subcontinent, while different geographical landscapes within other tiger subspecies maybe genetically less differentiated (Malayan peninsula). Our results contradict suggestions of population structure in wild Amur tigers (Sorokin et al., 2016), substantiate the significance of structure in Bengal tigers, and uncover hitherto unknown structure in tigers from the Malayan peninsula.

Which tigers retain the most variation?

Our data and analyses reveal Bengal tigers have the highest genetic variation across the genome. This is to be expected given historical records of tiger occupancy (Karanth et al., 2010, across a large variety of habitats and subsisting on a wide range of prey species that range from the large rhinoceros and gaur to the small hog deer and barking deer), present population sizes of tigers in India, and previous genetic studies

based on a limited number of DNA microsatellite markers (Mondol et al., 2009). In contrast, only Bengal tigers reveal signatures of potentially recent inbreeding, indicating substantial isolation between populations. Simulations of demographic history suggested strong signatures of very intense and recent bottlenecks (modeled at around 50 generations ago) in Bengal tiger populations. High total genetic variation accompanied by recent inbreeding is reflective of the intense effect of hunting in India just a century ago (Rangarajan 2006), followed by extensive habitat loss and ongoing isolation of populations. In comparison, Amur individuals do not harbor long homozygous stretches in their genomes, while individual Bengal tigers do. A closer look at landscapes and habitats in India and the Russian Far East reveal strong differences: India is dominated by variable habitats amidst a matrix of extremely high human population densities, while in the Russian Far East, human density is low and habitat is more continuous (Miquelle et al., 2010). Indeed, landscape genetics studies have suggested that high human population density is a barrier for tiger movement (Thatte et al., 2018). We suggest that extreme fragmentation and high human population density in India has resulted in isolated populations, where individuals may be more likely to mate with relatives. In contrast, despite low Amur tiger population densities in the Russian Far East, individual movement is not hindered by significant barriers and the population is more panmictic, with little to no sign of geographic population substructure (e.g. see Henry et al., 2010).

Within Bengal tigers, we observe high variance in long ROH, generally thought to be the consequence of recent inbreeding, or possibly recent, intense bottlenecks. For example, tigers from Central India retain lower proportions of long ROH than those from

other Indian landscapes (e.g. Western India, South India), possibly an outcome of higher recent connectivity between the Central Indian tiger populations (Thatte et al. 2018), or lower historical bottleneck intensity. These very specific and hierarchical results underscore the importance of the inclusion of multiple genome-wide sampling across and within regions, as single representatives may be a poor reflection of inbreeding and variation for any given population, and do not provide a context with which to evaluate significance across subspecies and populations. In the future, simulations that incorporate realistic recombination rates could be used to model and disentangle the cumulative impacts of recent demographic history and very recent inbreeding on distributions of ROH in the genome.

What evolutionary processes dominate the evolution of tigers and their subspecies?

Our models and analyses suggested relatively recent divergence between tiger populations (last 9,000 or so years versus 68,000 years inferred by Liu et al., 2018), highlighting the role of drift/stochastic processes in recent tiger evolution. Our analyses are based on several genomes per populations, whereas the G-PhoCS analyses in Liu et al., (2019) a single genome from each population. We observed that shorter ROH (potentially indicative of historical bottlenecks) are well represented across all individuals, consistent with our demographic inferences from the coalescent simulations of intense historical bottlenecks. Our results consistently underline the genome-wide importance of genetic drift. Because populations have differentiated recently, we might not expect to find significant genetic differences. However, because all tiger populations have been through intense bottlenecks following divergence, we observed strong

signals of population differentiation.

The order of divergence of the subspecies from the ancestral tiger metapopulation is partially consistent with previous suggestions of tigers being isolated in Sumatra first, likely due to sea level rise (consistent with previous research, e.g. Liu et al., 2018), closely followed in time by those in India, then last by populations in northeast Asia and Malaysia (not consistent with Liu et al., 2018). Within Bengal tigers, we observed that northeast Indian tigers diverged considerably earlier than other Bengal populations. We re-iterate that although northeast Indian tigers are the most distinct of Bengal tigers, they are closer to Bengal tigers than they are to any other tigers. The northeast Indian tigers in this study are from the state of Assam, and sampling other, more eastern populations from this remote region might yield interesting insights, as would samples from the Indo-Chinese tigers.

Our results suggested that Amur tiger genomes demonstrate signals of selection, with possible adaptations to a colder environment. These results are consistent with recent studies in two human populations that live in cold environments, including Greenlandic Inuit (Fumagalli et al., 2015) and Indigenous Siberians (Hallmark et al., 2018), which revealed signatures of selection on genes and pathways involved in lipid metabolism. Europeans also reveal over-representation of Neanderthal mutations involved in lipid catabolism (Khrameeva et al., 2014). Similarly, polar bears genomes also reveal signatures of selection on lipid metabolism genes (Liu et al., 2014). Understanding the distribution of adaptive variants could be important for future conservation efforts, especially if priority were placed on preserving these cold-adapted populations.

Sumatran tigers appear to have experienced strong genetic drift following vicariance from mainland south east Asia, maintained a smaller effective population size, and have experienced a strong recent bottleneck. Sumatran tigers also appear to harbour some signatures of selection in their genomes. Liu et al., 2018 suggested selection for body size that targeted the *ADH7* gene, we did not detect any signature of selection at this locus in our Sumatran samples. However, we identify alternative candidate genes that can be involved in body size, such as the genes found in the *Regulation of anatomical structure size* GO term (Table 1). Difference in loci identified under selection between Lui et al. and our study maybe because we ascertained for loci under selection contingent on our estimated demographic history, while Liu et al. (2019) used genome scans that did not incorporate demographic history. We caution that it is difficult to truly distinguish among all population genetic processes, especially selection, without more data, and assignments of GO categories designed from model organisms are only a substitute for more definitive tests of selection.

We did not detect signatures of selection or extensive gene flow into Malayan and Bengal tiger genomes, suggesting genomic variation impacted primarily by drift. Indian tigers appear to have experienced intense founding events, intense recent bottlenecks, and population structuring, suggesting a relatively stronger role for drift (compared with Malayan tigers) in shaping genome-wide variation.

Conservation implications

Our analyses suggest that tigers have recently differentiated into subspecies with unique gene pools, and contrasting histories of drift and selection make each of these

four putative subspecies evolutionarily unique. We present preliminary information on selection that warns of potentially serious negative consequences of exchanging individuals between populations of separate subspecies. Our analyses suggest that introgression from other gene pools into Amur and Sumatran tigers could result in outbreeding depression and/or loss of their unique adaptations. Similarly, using source populations from a different subspecies for reintroduction efforts, as has been proposed (Launay et al. 2012), may have unintended consequences. Conservation has mostly relied on reconstruction of history as a guide to management of threatened species, however it may be important to consider novel environments in a changing world. Poleward progression of subtropical climates may favor adaptive alleles (found in more southern Sumatran tigers) in more northern populations (e.g. Malayan or Thai tigers), potentially increasing the value of these adaptations to future survival.

Active exchange of individuals among selected sub-populations of Bengal tigers, as has been done for the Florida panther (Johnson et al. 2010) may become a critical management tool given the potential negative impacts of inbreeding and drift in these populations. Within India, it is critical that the management status of northeast Indian tigers be re-evaluated given our results on their antiquity. Ongoing human impacts like fragmentation will likely continue to disrupt natural evolutionary processes. Managing local populations to minimize human impacts and allowing continued tiger evolution may be the key to species survival, and the important conservation strategy for the Anthropocene.

Materials and Methods

Sample Collection

We obtained tissue, blood, or serum samples from as many geographically distinct tiger populations as possible. This amounted to 66 samples from 4 tiger subspecies including 21 Indian Bengal tigers (*P. t. tigris*), 19 Eastern Siberian tigers (*P. t. altaica*) (including the published individual in NCBI, SRA ID: SRX272997), 17 Malayan tigers (*P. t. jacksoni*), and 11 Sumatran tigers (*P. t. sumatrae*). A list of final samples sequenced and their sources are available in Supplementary Table 1.

Reference assembly sequencing and de novo assembly

We received whole blood from a wild born Malayan tiger (*P. t. jacksoni*) sampled by the El Paso zoo, Texas on 7/28/2016, collected as part of a routine health checkup. We immediately froze the sample at -80ºC until it was shipped on dry ice to the Barsh lab at HudsonAlpha for extraction and delivery to the Genome Services Lab (GSL) at HudsonAlpha Institute for Biotechnology, Huntsville, Alabama. DNA was extracted and purified using the Qiagen MagAttract HMW DNA kit. GSL staff prepared a linked-read sequencing library using the Chromium controller. The library was sequenced on one lane of a HiSeqX. We assembled the genome using the SuperNova assembly software (1.1.4) provided by 10x Genomics using the standard pipeline. We refer to this assembly as Maltig1.0 hereafter.

Whole Genome Re-sequencing

We extracted DNA from samples using the Qiagen DNeasy blood and tissue kits (Catalogue #69504) and quantified DNA concentrations with the Qubit dsDNA HS assay

kit (Q32851). As a number of our samples yielded very low amounts of DNA (< 1 ng), we used an approach that scales down the input reagents from an Illumina Nextera kit (Baym et al. 2015). Genomic libraries were run on a 2100 BioAnalyzer (Agilent Technologies using High Sensitivity DNA Chips (Catalog #5067-4626) to determine the quality, quantity, and fragment size distributions. Libraries were sequenced on Illumina HiSeqX, HiSeq 4000, and HiSeq 2500 for between 5-25x coverage using paired-end 2x150bp reads (Supplementary Table 1).

Variant Discovery

We trimmed reads prior to mapping with TrimGalore (Krueger 2015), then mapped reads to the Maltig1.0 reference genome using BWA-MEM (Li 2013) and sorted and indexed using SAMtools (Li et al. 2009). We marked duplicate reads with the Picard Tools `MarkDuplicates` command (http://broadinstitute.github.io/picard). We then called variants from the resulting BAM files using FreeBayes (Garrison & Marth 2012). The resulting VCF file was filtered with VCFtools (Danecek et al. 2011) to a minimum quality of 30, a genotype quality of 30, maximum of 2 alleles, Hardy-Weingburg p-value of 0.0001, minimum allele frequency of 0.01, and a minimum minor allele count of 3. The pipeline was managed and parallelized using NextFlow (Tommaso et al. 2017). The full pipeline scripts and commands are available in the supplemental materials.

Population structure

We first investigated admixture and structure between populations using Plink2 (Chang et al. 2015). We used VCFtools to filter the initial variant call file using ‘max-

missing 0.95' and 'maf 0.025' to remove sites with missing data and rare variant calls. We then converted to Plink's '.ped/.map' format using VCFtools, and subsequently converted to '.bed/.bim/.fam' format within Plink2 using the flag '--make-bed'. PCA was then run on the resulting bed file using the flag '--pca 10' which computes the variance-standardized relationship matrix. PCAs were then plotted using R. For smaller runs an additional step was added within Plink2 to first calculate the frequencies using the flag '--freq'. Subsequently, PCA was run using the '--pca' flag and inputting the frequency file using the '--read-freq' flag. We used this protocol on the vcf with all individuals and subsequently, we divided the vcf into the putative subspecies for within subspecies runs.

The program ADMIXTURE was used to infer structure between populations and inform clusters which represent populations with distinct histories (Alexander et al. 2009). ADMIXTURE uses maximum likelihood-based models to infer underlying ancestry for unrelated individuals. We used the filtered dataset (VCFools max-missingness cutoff of 95%, minor allele frequency cutoff of 0.025) and resulting Plink formatted files for input into the software. In order to infer the most likely value of K, values of 2-8 were run. We also performed K validation in order to compute the cross-validation error for each value of K, by using the '--cv' flag within the program. The value with the least error is informative of the best value of K for the data.

The program NetStruct_Hierarchy was used to construct a population structure tree (PST), representing hierarchical population structure (Greenbaum et al. 2019). The genetic-similarity network was constructed from the same data used for ADMIXTURE and PCA. To construct the PST, edge-pruning of the network was conducted by

incrementing the edge-pruning threshold by increments of 0.0001, and conducting network cluster detection until reaching clusters of size 3 (the smallest possible cluster size).

Rarefaction analysis

To ensure that our data was reflective of the diversity within each subspecies/unit as defined by ADMIXTURE, we used the program ADZE (Szpiech et al. 2008). ADZE runs a rarefaction analyses on polymorphism data in order to estimate the number of alleles private to any given population (not found in any other population), considering equal-sized subsamples from each input population. In addition, the program calculates distinct alleles within each population. We calculated the private alleles across the four main populations/sub-species as designated by the ADMXITURE software, in addition to the distinct alleles within each of the four populations individually.

Population differentiation and diversity

We calculated pairwise F_{ST} between each subspecies group as defined by ADMIXTURE using VCFtools. Variant call data was subdivided into sub-species based on PCA (Bengal, Sumatran, Amur, and Malayan as subgroups) and was used to compute pairwise F_{ST} between each group. Nucleotide diversity (π) was calculated using VCFtools.

In order to detect the number of single nucleotide variants (SNV), the data were filtered using VCFtools (Danecek et al. 2011) to a minimum base quality of 30, genotype quality of 30 and depth 10. We additionally filtered for minor allele frequency of 0.025

and allowed a maximum 5% missing data in any loci. RTG tools (https://www.realtimegenomics.com/products/rtg-tools) vcfstats was used to calculate the total number of heterozygous SNP sites for each individual. These values were then plotted alongside comparable estimates for other species reported in Abascal et al (2016).

Ancient demographic history

Pairwise sequentially Markovian Coalescent (PSMC) (Li and Durbin, 2011) is a single genome method to detect historical effective population size. In order to estimate historical population size changes for the different subspecies, we removed sex chromosome scaffolds for AMU1, MAL1, SUM2 and BEN_SI3 (the highest coverage individual for each subspecies). The procedures for sex chromosome filtering can be found in the supplementary text. Additionally, sites with a minimum of half the average sequencing depth or twice the average sequencing depth were filtered out while calling variant sites. The resulting scaffolds were then used to estimate the effective population size across 34 time intervals as described in Li and Durbin (2011). 100 rounds of bootstrap replicates were performed.

Demographic history with SFS and coalescent models

Demographic models

Data filtering procedures for the demographic models can be found in the Supplementary text. Using the program fastsimcoal 2 (Excoffier et al. 2013), we performed demographic estimations of the model shown in Figure 5 on two datasets in

two consecutive steps, such as to reduce the number of parameters to estimate simultaneously. The first step consisted in estimating the demography (24 parameters) of four tiger subspecies (Malaysia – MAL, Sumatra –SUM, Bengal – BEN, and Amur – AMU) using the individuals of each subspecies that had the highest coverage. We thus selected 3 SUM individuals, 3 BEN individuals from South India (BEN_SI), four MAL individuals, and 3 AMU individuals, which all had >20X coverage on average (See list in Supplementary Table 3). We modeled the four-subspecies as belonging to a large Asian metapopulation, from which they would have diverged some time ago while still receiving some continuous gene flow from the metapopulation. Note that this continent-island population structure amount to modeling a set of populations having gone through a range expansion (Excoffier 2004). We assumed that each of the four subspecies could have gone through two distinct bottlenecks, one that would have occurred at the time of the separation from the Asian metapopulation, to mimic some initial founder effect, and one that would be recent to mimic habitat deterioration. We also assumed that the Asian metapopulation could have gone through an ancestral bottleneck sometime in the past.

The second step used estimated parameters in a more complex model including the specific demography of four Bengal tiger populations (24 new additional parameters). To this aim, as in the previous analysis, we selected individuals with the highest coverage (>20X) from each population (see Supplementary Table 1, samples used represented in Supplementary Table 4). No individuals from BEN_NOR were included as their coverage was low and they are part of the same genetic cluster as BEN_CI. In order to correctly estimate the relationship between these populations and

the other subspecies, we also included 3 MAL individuals in this analysis. The new model included all the parameters from the previous model, fixed at their previously estimated values, except some parameters re-estimated for the BEN_SI population, which was now assumed to have diverged from an Indian metapopulation at some time in the past, like the other three BEN populations. We also estimated the size and the divergence of the BEN metapopulation from the Asian metapopulation. We allowed the sampled BEN populations to have gone through two bottlenecks (an initial founder effect and a recent bottleneck). The parameter estimated in these two steps are shown in Supplementary Tables 6 and 7) and the resulting demography is sketched in Figure 3.

Parameter estimation and fastsimcoal2 command line

Fastsimcoal2 (Excoffier et al. 2013) was used to estimate parameters from the multidimensional site frequency spectrum (SFS) computed on non-coding regions at least 50 Kb away from known genes. The multidimensional SFS was computed with the program Arlequin ver 3.5.2.2 (http://cmpg.unibe.ch/software/arlequin35) on polymorphic sites matching the filtering criteria listed above. In order to infer absolute values of the parameters, we used a mutation rate of 0.35e-8 per site per generation estimated in a previous paper on tiger demography (Liu et al. 2018), and we assumed that the proportion of monomorphic sites passing our filtering criteria in each 1Mb segment was identical to the fraction observed in polymorphic sites. Parameter estimation was obtained by maximum-likelihood estimation obtained from 100 independent runs of

fastsimcoal2, 60 expectation conditional maximization (ECM) cycles per run and 500 thousand coalescent simulations per estimation of expected SFS. The *fastsimcoal2* command line used for the estimation was of the type

```
fsc -t xxx.tpl -n500000 -d -e xxx.est -M -l30 -L60 -q  -C1 --multiSFS -c1 -B1
```

where fsc is the *fastsimcoal2* program and xxx the generic name of the input files. The .est and .tpl input files used for inference on the two datasets are listed in the Supplementary material.

Confidence intervals were estimated via a block bootstrap approach. We generated 100 bootstrapped SFS by resampling (and adding) the SFS from 1 Mb segments of concatenated non-coding segments along the genome, and for each of these bootstrapped samples we re-estimated the parameters of the model using 10 *fastsimcoal2* independent runs starting at the maximum-likelihood parameter values. Again, we used 60 ECM cycles for each run and performed 500,000 simulations for estimating the expected SFS under a given set of parameter values. The limits of 95% confidence intervals were estimated by computing the 2.5% and 97.5% quantiles of the distribution of 100 maximum-likelihood parameter values.

Genome scan for selection

To detect the footprints of natural selection in different tiger subspecies, we analyzed individuals with coverage > 10X from 4 subspecies (n = 34). We filtered out genotypes with depth of coverage < 10 (DP) and genotype quality < 30 (GQ). We excluded scaffolds shorter than 1 Mb. We kept sites with no missing data among the 34 individuals.

We considered the 4 subspecies as 4 populations and computed pairwise F_{ST} values (Hudson et al. 1992) along the genome over 50 kb sliding windows (with a step of 10 kb) using the R package PopGenome (Pfeifer et al. 2014). We then computed a measure of selection similar to the Population Branch Statistic (PBS) (Yi et al. 2010, Shriver et al. 2004). The PBS statistic is based on a three-population comparison and measures the length of the branch leading to a given population since its divergence from the two other populations. This statistic is not able to accommodate more than three populations and relies on a tree-based model that does not correspond to tigers' demographic history. Therefore, we extended this statistic to the case of four populations under a more suitable model than a tree-based one. Furthermore, using all four populations allows to better characterize the differences that are exclusive to specific branch. We define:

$$mPBS_{\mathrm{a}} = \frac{2(T_{\mathrm{ab}}+T_{\mathrm{ac}}+T_{\mathrm{ad}})-(T_{\mathrm{bc}}+T_{\mathrm{bd}}+T_{\mathrm{cd}})}{6},$$

where T_{ij} is the divergence time, in generations, between population *i* and *j* (Nei, 1972):

$$T_{ij} = -\log(1 - F_{\mathrm{ST}}^{ij}).$$

This statistic assumes that each population diverged from a metapopulation independently and that no migration occurred following divergence. It measures the length of the branch leading to a given lineage since its divergence from the metapopulation. Selection in a given lineage will lead to a much longer terminal branch than under neutrality. This would translate to extreme mPBS values.

To compare observed mPBS values to expectations under the tigers' demographic history, we simulated 1 million genomic windows using the demographic

model inferred previously. Window size and sample sizes for each population are the same as in the observed dataset. Parameter values are fixed and correspond to the maximum likelihood estimates (Table S7). Then, we computed the mPBS statistic for each population to generate a null distribution. Observed and simulated distributions were then represented to see whether observed values deviated from neutral expectations.

To identify putative genes under selection, we considered predicted genic regions of the tiger genome for which a homolog has been annotated using Exonerate (protein2genome model). To avoid spurious enrichment signals due to the presence of multiple homologs for a single gene, we kept only one homolog for each predicted gene. If different homologs on the same strand overlap, we pick the first one and ignore the others. We retained a total of 12,771 genes after filtering.

We also checked whether some Gene Ontology (GO) terms (Ashburner et al. 2000, Mi et al. 2016) were enriched across candidate genes (Fisher's exact test performed on human GO terms). Genes (+/- 50 kilobases flanking regions) were considered as candidates if they overlapped with a window that was in the top 0.1 % of mPBS value of a given population. The reference list of genes for the enrichment test is set as the list of genes after filtering (12,771 genes).

Runs of Homozygosity

To estimate runs of homozygosity (ROH), we used the filtered SNPs from the autosomal scaffolds. Individuals with more than 10x average coverage were grouped as per subspecies. We used BCFtools/RoH (Narasimhan et al 2016) to estimate ROH. The

autozygous runs obtained were classified into various lengths (runs between 10kb and 100kb, runs between 100kb and 1 Mb, and runs longer than 1 Mb). Proportion of genome in ROH for 1Mb was estimated as the total length of the genome in more than 1Mb runs divided by the total length of autosomal scaffolds. Similar calculations were made for 100kb to 1Mb runs, and for 10kb-100kb runs except the length of the genome longer than 1Mb and 100Kb were subtracted from total length of autosomes respectively.

Estimation of ROH using sliding window approach

Within individual genomes, we identified ROH using the methods first implemented in Pemberton et al. (2012). For each of the four tiger populations, we estimated the allele frequencies at each SNP using the observed allele frequencies for the four populations in our data set. To identify ROH, we employed a likelihood method from Wang et al. (2009) adapted by Pemberton et al. (2012). This approach, which forms the basis of the ROH inference program GARLIC (Szpiech et al. 2017), considers a sliding window of *n* SNPs that moves along the chromosome in single SNP increments. To ensure robustness in our results, we repeated the ROH identification procedure with three values of *n*: 100, 250, and 700. Within each window, a log-likelihood score was computed for each SNP, comparing the hypothesis that the SNP is autozygous to the hypothesis that it is non-autozygous, allowing for an error term that accounts for mutation or genotyping error. As in Pemberton et al. (2012), we set the error parameter to 0.001. The overall score of a window is then the sum of the scores of the SNPs it contains, with an observed homozygous SNP contributing a positive score

(unless the SNP is monoallelic in the population in question), and an observed heterozygous SNP contributing a negative score (Wang et al., 2009).

Following that, all windows with an overall score of 0 or greater were taken to be in an ROH, with overlapping windows merged and considered as part of a single ROH. The bp length of each ROH is taken to be the length of the interval between the its two most extreme SNPs, including the endpoints. For the three values of *n*, we then plotted the total length of ROH present in each individual, as well as the total length of long ROH (≥ 1Mb) present in each individual (Supplementary Figure 13).

Acknowledgements

We thank the American Zoo Association for endorsing our research and collection of samples from captive tigers, Tara Harris (then Minnesota Zoo) and Kathy Traylor-Holzer (Tiger Species survival program) for help with captive tiger studbooks. We thank San Francisco Zoo, San Diego Zoo (BR2016035; Leona Chemnick for assistance with DNA extraction and sample transport), El Paso zoo, Omaha Zoological Society, WCS Bronx zoo (IC2016-0464 WCS; Dee McAloose and Jean Pare for assistance with sample transport), Gopala Battu for assistance with sequencing and sample preparation at Hudson Alpha. Zachary Szpiech for assistance with ADZE. U Ramakrishnan thanks National Tiger Conservation Authority and R Gopal for samples from Ranthambore. U Borthakur thanks Assam Forest Department. YV Jhala thanks the Chief Wildlife Wardens of Uttarakhand, Madhya Pradesh and Rajasthan, and the ministry of Environment and Forests for permissions to radio-collar tigers and collect blood samples. Support was provided by Infosys Travel Award to A Khan, SciGenome

Research Foundation grant to A Khan, Fulbright Nehru Academic exchange fellowship to U Ramakrishnan, CEHG, Stanford University funding to U Ramakrishnan, Wellcome Trust-DBT Indian Alliance Senior fellowship to U Ramakrishnan (IA/S/16/2/502714), Genomics Facility of CCAMP. We thank Atul Upadhayay for bioinformatics support, and the computing facility at NCBS.

Author contributions: EA, AK, RWT, UR, and EAH designed the study. EA, AK, RWT, AG, GG, AT, JTLK, SR, SP, GB, CK, SP, AC, LE conducted lab work and analyses. GB, CK, MG, DM, AZ, UB, AR, EL, OAR, YVJ, EAH, UR provided samples and funding. EA, AK, RWT, AG, GG, AT, JTLK, SR, SP, GB, CK, SP, AC, MG, DM, AZ, UB, AR, EL, OAR, YVJ, DP, LE, EAH, UR wrote and edited the paper.

References

1. Abascal F, Corvelo A, Cruz F, Villanueva-Cañas JL, Vlasova A, Marcet-Houben M, et al. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biol.* 2016 Dec 14;17(1):251.

2. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009 Sep;19(9):1655–64.

3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Michael Cherry J, et al. Gene Ontology: tool for the unification of biology [Internet]. Vol. 25, *Nature Genetics.* 2000. p. 25–9. Available from: http://dx.doi.org/10.1038/75556

4. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One.* 2015 May 22;10(5):e0128036.

5. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience.* 2013 Jul 22;2(1):10.

6. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 2018 Apr;19(4):220–34.

7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015 Feb 25;4:7.

8. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, et al. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nat Commun.* 2013;4:2433.

9. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools [Internet]. Vol. 27, *Bioinformatics.* 2011. p. 2156–8. Available from: http://dx.doi.org/10.1093/bioinformatics/btr330

10. Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, et al. Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol.* 2013 Jul;30(7):1544–58.

11. Ellegren H, Galtier N. Determinants of genetic diversity. *Nat Rev Genet.* 2016 Jul;17(7):422–33.

12. Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, Schönfeld B, Wiench C, Hohenlohe P, Storfer A. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nat Commun.* 2016 Aug 30;7:12684.

13. Excoffier L. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol.* 2004. Vol. 13:853–64.

14. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 2013 Oct;9(10):e1003905.

15. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science.* 2015 Sep 18;349(6254):1343–7.

16. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* [q-bio.GN]. 2012. Available from: http://arxiv.org/abs/1207.3907

17. Goodrich, J., Lynam, A., Miquelle, D., Wibisono, H., Kawanishi, K., Pattanavibool, A., Htun, S., Tempa, T., Karki, J., Jhala, Y. & Karanth, U. 2015. *Panthera tigris*. *The IUCN Red List of Threatened Species* 2015:

e.T15955A50659951. http://dx.doi.org/10.2305/IUCN.UK.2015-2.RLTS.T15955A50659951

18. Gouy A, Daub JT, Excoffier L. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Res.* 2017 Sep 19;45(16):e149.

19. Greenbaum G, Rubin A, Templeton AR, Rosenberg NA. Network-based hierarchical population structure analysis for large genomic datasets. *bioRxiv*. 2019 [cited 2019 May 25]. p. 518696. Available from: https://www.biorxiv.org/content/10.1101/518696v1

20. Hallmark B, Karafet TM, Hsieh P, Osipova LP, Watkins JC, Hammer MF. Genomic Evidence of Local Adaptation to Climate and Diet in Indigenous Siberians. *Mol Biol Evol.* 2019 Feb 1;36(2):315–27.

21. Heaney LR. A synopsis of climatic and vegetational change in Southeast Asia. In Tropical Forests and Climate 1991 (pp. 53-61). Springer, Dordrecht.

22. Henry, P., D. Miquelle, T. Sugimoto, D. R. McCullough, A. Caccone, M. A. Russello. 2009. In situ population structure and ex situ representation of the endangered Amur tiger. *Mol Ecol.* 18:3173–3184.

23. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. *Genetics.* 1992 Oct;132(2):583–9.

24. Jhala YV, Qureshi Q, Gopal R, Sinha PR. The Status of Tigers in India 2014. National Tiger Conservation Authority.

25. Johnson, W. E., D. P. Onorato, M. E. Roelke, E. D. Land, M. Cunningham, R. C. Belden, R. McBride, D. Jansen, M. Lotz, D. Shindle, J. Howard, D. E. Wildt, L. M.

Penfold, J. a Hostetler, M. K. Oli, and S. J. O'Brien. 2010. Genetic restoration of the Florida panther. *Science* 329:1641–1645.

26. Joshi A, Vaidyanathan S, Mondol S, Edgaonkar A, Ramakrishnan U. Connectivity of Tiger (Panthera tigris) Populations in the Human-Influenced Forest Mosaic of Central India. *PLoS One*. 2013. Vol. 8, p. e77980.

27. Joshi AR, Dinerstein E, Wikramanayake E, Anderson ML, Olson D, Jones BS, Seidensticker J, Lumpkin S, Hansen MC, Sizer NC, Davis CL. Tracking changes and preventing loss in critical tiger habitat. *Science advances.* 2016 Apr 1;2(4):e1501675.

28. Karanth KK, Nichols JD, Karanth KU, Hines JE, Christensen NL Jr. The shrinking ark: patterns of large mammal extinctions in India. *Proc Roy Soc B* 2010 Jul 7;277(1690):1971–9.

29. Kardos M, Åkesson M, Fountain T, Flagstad Ø, Liberg O, Olason P, et al. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol.* 2018 Jan;2(1):124–31.

30. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PLoS One.* 2010 Nov 15;5(11):e13996.

31. Kitchener, A, Breitenmoser, C, Eizirik, E, Gentry, A, Werdelin, L, Wilting, A, Yamaguchi, N, Abramov, A, Christiansen, P, Driscoll, C, Duckworth, W, Johnson, W, Luo, S, Meijaard, E, O'Donoghue, P, Sanderson, J, Seymour, K, Bruford, M, Groves, C, Tobe, S (2017). A revised taxonomy of the Felidae. The final report of

the Cat Classification Task Force of the IUCN/SSC Cat Specialist Group. Cat News Special Issue. 80 pp..

32. Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015;

33. Launay, F., N. Cox, M. Baltzer, T. Tepe, J. Seidensticker, R. Krishnamurthy, T. Gray, S. Christie, S. Simcharoen, R. Singh, S. Lumpkin, C. Bruce, and M. Wright. 2012. Preliminary Study of the Feasibility of a Tiger Restoration Programme in Cambodia's Eastern Plains: A Report Commissioned by World Wide Fund for Nature.

34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009 Aug 15;25(16):2078-9.

35. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011 Jul 13;475(7357):493–6.

36. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. *arXiv* [q-bio.GN]. 2013. Available from: http://arxiv.org/abs/1303.3997

37. Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell.* 2014 May 8;157(4):785–94.

38. Liu Y-C, Sun X, Driscoll C, Miquelle DG, Xu X, Martelli P, et al. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Curr Biol.* 2018;28(23):3840–9.

39. Luo S-J, Kim J-H, Johnson WE, van der Walt J, Martenson J, Yuhki N, et al. Phylogeography and genetic ancestry of tigers (Panthera tigris). *PLoS Biol.* 2004 Dec;2(12):e442.

40. Luo, S. J., Liu, Y. C., & Xu, X. (2019). Tigers of the World: Genomics and Conservation. *Ann rev of animal biosc*, *7*, 521-548.

41. Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, et al. Quantification of extinction risk: IUCN's system for classifying threatened species. *Conserv Biol.* 2008 Dec;22(6):1424–42.

42. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D183–9.

43. Miquelle, D. G., J. M. Goodrich, E. N. Smirnov, A. Phillip, O. Y. Zaumyslova, G. Chapron, L. Kerley, A. A. Murzin, M. G. Hornocker, and H. B. Quigley. 2010. Amur tiger: a case study of living on the edge. "Conservation of Wild Felids". Oxford University Press, Oxford.

44. Mohr DW, Naguib A, Weisenfeld N, Kumar V, Shah P, Church DM, et al. Improved de novo Genome Assembly: Linked-Read Sequencing Combined with Optical Mapping Produce a High Quality Mammalian Genome at Relatively Low Cost. *bioRxiv.* 2017 [cited 2019 Mar 18]. p. 128348. Available from: https://www.biorxiv.org/content/10.1101/128348v2.abstract

45. Mondol S, Ullas Karanth K, Ramakrishnan U. Why the Indian Subcontinent Holds the Key to Global Tiger Recovery. *PLoS Genetics.* 2009. Vol. 5, p. e1000585. Available from: http://dx.doi.org/10.1371/journal.pgen.1000585

46. Murchison EP, Schulz-Trieglaff OB, Ning Z, Alexandrov LB, Bauer MJ, Fu B, et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell.* 2012 Feb 17;148(4):780–91.

47. Murray GGR, Soares AER, Novak BJ, Schaefer NK, Cahill JA, Baker AJ, et al. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science.* 2017 Nov 17;358(6365):951–4.

48. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics.* 2016 Jun 1;32(11):1749–51.

49. Natesh M, Atla G, Nigam P, Jhala YV, Zachariah A, Borthakur U, et al. Conservation priorities for endangered Indian tigers through a genomic lens. *Sci Rep.* 2017 Aug 29;7(1):9614.

50. Nei, M. (1972). Genetic distance between populations. *The Amer Nat*, *106*(949), 283-292.

51. Nowell K, Jackson P (1996) Wild cats: Status survey and conservation action plan. Gland, Switzerland: IUCN-World Conservation Union. 406 p.

52. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol.* 2015 May 18;25(10):1395–400.

53. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet.* 2012 Aug 10;91(2):275–92.

54. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 2014 Jul;31(7):1929–36.

55. Pool JE, Hellmann I, Jensen JD, Nielsen R. Population genetic inference from genomic sequence variation. *Genome research.* 2010 Mar 1;20(3):291-300.

56. Rangarajan M (2006) India's Wildlife History: an Introduction. Orient Longman.

57. Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, Wayne RK. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances.* 2019 May 1;5(5):eaau0757.

58. Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I. Inbreeding and extinction in a butterfly metapopulation. Nature. 1998 Apr;392(6675):491.

59. Singh R, Qureshi Q, Sankar K, Krausman PR, Goyal SP. Use of camera traps to determine dispersal of tigers in semi-arid landscape, western India. *J Arid Environ.* 2013 Nov 1;98:105–8.

60. Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, et al. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics.* 2004 May;1(4):274–86.

61. Sorokin PA, Rozhnov VV, Krasnenko AU, Lukarevskiy VS, Naidenko SV, Hernandez-Blanco JA. Genetic structure of the Amur tiger (Panthera tigris altaica) population: Are tigers in Sikhote-Alin and southwest Primorye truly isolated? *Integr Zool.* 2016;11(1):25–32.

62. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics.* 2008 Sep 8;24(21):2498-504

63. Sutherland GD, Harestad AS, Price K, Lertzman KP. Scaling of natal dispersal distances in terrestrial birds and mammals. *Conserv Ecol.*2000;4(1). Available from: https://www.jstor.org/stable/26271738

64. Thapa K, Manandhar S, Bista M, Shakya J, Sah G, Dhakal M, et al. Assessment of genetic diversity, population structure, and gene flow of tigers (Panthera tigris tigris) across Nepal's Terai Arc Landscape. *PLoS One*. 2018 Mar 21;13(3):e0193495.

65. Thatte P, Joshi A, Vaidyanathan S, Landguth E, Ramakrishnan U. Maintaining tiger connectivity and minimizing extinction into the next century: Insights from landscape genetics and spatially-explicit simulations. *Biol Conserv.* 2018 Feb 1;218:181–91.

66. Tommaso DP, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol.* 2017 Apr 11;35(4):316–9.

67. Wang S, Haynes C, Barany F, Ott J. Genome-wide autozygosity mapping in human populations. *Genet Epidemiol.* 2009 Feb;33(2):172–80.

68. Wilting A, Courtiol A, Christiansen P, Niedballa J, Scharf AK, Orlando L, et al. Planning tiger recovery: Understanding intraspecific variation for effective conservation. *Science Advances.* 2015. Vol. 1, p.e1400175.

69. Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science.* 2015. Vol. 348, p. 242–5.

70. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* 2010 Jul 2;329(5987):75–8.

71. Yumnam, B., Jhala, Y. V., Qureshi, Q., Maldonado, J. E., Gopal, R., Saini, S., ... & Fleischer, R. C. (2014). Prioritizing tiger conservation through landscape genetics and habitat linkages. *PLoS One*, *9*(11), e111207.