PNAS RESEARCH ARTICLE | POPULATION BIOLOGY

Unraveling the genomic diversity and admixture history of captive tigers in the United States

Ellie E. Armstrong[a,1,2], Jazlyn A. Mooney[a,b,1,2], Katherine A. Solari[a], Bernard Y. Kim[a], Gregory S. Barsh[c,d], Victoria B. Grant[a], Gili Greenbaum[e], Christopher B. Kaelin[d], Katya Panchenko[a], Joseph K. Pickrell[f], Noah Rosenberg[a], Oliver A. Ryder[g], Tsuya Yokoyama[a], Uma Ramakrishnan[h], Dmitri A. Petrov[a,i,j], and Elizabeth A. Hadly[a,k,l,m,2]

Affiliations are included on p. 10.

Edited by Marcus Feldman, Stanford University, Stanford, CA; received February 20, 2024; accepted August 9, 2024

Genomic studies of endangered species have primarily focused on describing diversity patterns and resolving phylogenetic relationships, with the overarching goal of informing conservation efforts. However, few studies have investigated genomic diversity housed in captive populations. For tigers (*Panthera tigris*), captive individuals vastly outnumber those in the wild, but their diversity remains largely unexplored. Privately owned captive tiger populations have remained an enigma in the conservation community, with some believing that these individuals are severely inbred, while others believe they may be a source of now-extinct diversity. Here, we present a large-scale genetic study of the private (non-zoo) captive tiger population in the United States, also known as "Generic" tigers. We find that the Generic tiger population has an admixture fingerprint comprising all six extant wild tiger subspecies. Of the 138 Generic individuals sequenced for the purpose of this study, no individual had ancestry from only one subspecies. We show that the Generic tiger population has a comparable amount of genetic diversity relative to most wild subspecies, few private variants, and fewer deleterious mutations. We observe inbreeding coefficients similar to wild populations, although there are some individuals within both the Generic and wild populations that are substantially inbred. Additionally, we develop a reference panel for tigers that can be used with imputation to accurately distinguish individuals and assign ancestry with ultralow coverage (0.25×) data. By providing a cost-effective alternative to whole-genome sequencing (WGS), the reference panel provides a resource to assist in tiger conservation efforts for both ex- and in situ populations.

Panthera tigris | inbreeding | captive tigers | admixture

Significance

In this study, we use genomic sequencing to investigate the diversity of the privately owned captive tiger population ("Generic" tigers) in the United States. Privately owned captive populations of tigers vastly outnumber both wild and accredited zoo tiger populations, making them an important consideration for future conservation efforts. Our study reveals that the captive population is neither inbred nor highly diverse, despite containing ancestry from all six extant wild tiger subspecies. Our accompanying reference panel can rapidly assign ancestry and individually identify tigers using ultralow coverage genomic data, providing a cost-effective and computationally efficient alternative to other more expensive and time-consuming methods and can be used to support ex and in situ monitoring and management decisions for tigers

The Anthropocene has been characterized by population declines, isolation, and extinction, leading to a shift in the abundance and distribution of thousands of species on Earth (1). The tiger (*Panthera tigris*) is one of the most iconic and captivating terrestrial species on the planet and exemplifies the severe and rapid declines in population size and reduced connectivity that threatens the survival of many species (2). Although some tiger populations and subspecies have boasted recent recovery efforts, others have gone extinct in the wild (3) or have been completely extirpated (4) due to human pressures. Today, tigers are composed of six subspecies (*P. t. altaica*, Amur; *P. t. tigris*, Bengal; *P. t. corbetti*, Indochinese; *P. t. jacksoni*, Malayan; *P. t. amoyensis*, South China; *P. t. sumatrae*, Sumatran) that have been genetically and geographically separated for at least the last 10,000 y (5, 6). Their geographic ranges span the Russian Far East and northeast China (*P. t. altaica*), Bangladesh, Bhutan, India, and Nepal (*P. t. tigris*), Myanmar, Thailand, and Laos (*P. t. corbetti*), the island of Sumatra (*P. t. sumatrae*), and the Malay Peninsula (*P. t. jacksoni*), respectively, while the South China tiger (*P. t. amoyensis*) is extinct in the wild with the last sighting more than 30 y ago (7). All subspecies have undergone severe population bottlenecks over the last century, primarily due to human activities such as hunting and the expansion of agriculture, which have directly reduced tiger numbers, habitat availability, and prey (8) and significant conservation efforts have been directed at their preservation and recovery since they are considered an ecologically important umbrella species (3, 4).

In contrast to wild populations, the global captive tiger population is now estimated to include 15,000 to 20,000 individuals worldwide, a number that is at least five times larger than the wild population (7). Most populations are largely unregulated, such as farmed and other privately owned tigers, while others, such as those in accredited zoos, are bred with the intention of serving as a diversity reservoir for dwindling wild populations. In the United States, the Association of Zoos and Aquariums (AZA) manages several tiger populations as distinct subspecies, specifically the Amur (1950s-present), Sumatran (1950s-present),

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]E.E.A. and J.A.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: ellieearmstrong@gmail.com, jazlynmo@usc.edu, or hadly@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2402924121/-/DCSupplemental.

Published September 19, 2024.

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

Malayan (1980s-present), and for a time the "Bengal" (white tigers; 1960s to 2011) tiger subspecies.

The idea that captive populations may serve as diversity reservoirs is a pervasive message from zoos that serves to justify the continued breeding of captive populations (9). For example, the regulated Amur tiger captive population was established in the early 1950s and has since increased to over 1,400 individuals in captivity worldwide (AZA; 10). Previous work that investigated diversity levels of 12 microsatellites in the Amur captive population in North America found that the population may be a reservoir of genetic diversity that is now extinct in the wild (11). This notion suggests that captive breeding programs with effective management are able to maintain genetic diversity and avoid inbreeding depression despite the small number of founders (in the case of the Amur tigers, 29 males and 28 females; 10). Genomic perspectives on whether captive populations have truly served as reserves for genetic diversity are rare and the conclusions are inconsistent across species (12–17).

Though there are a large number of tigers in captivity, only a small fraction of these tigers (less than 1%) reside in regulated facilities, such as accredited zoos (7). Outside of regulated establishments, individuals and entities breed, own, and sell tigers both legally and illegally. The formation of the privately owned, captive population in the United States likely coincided with the establishment of zoos and circuses in the early 1900s, since these entities were known to exchange individuals (8), but ultimately the ancestry and establishment timing of the captive, privately owned population is unknown. Over time, the mission of AZA zoos began to run counter to circuses and other private ownership practices, but by the 2000s, estimates put the captive tiger population in the United States at greater than 5,000 individuals (8, 18). Many of the unregulated facilities do not report basic or reliable population data and have emerged as a concern globally with the illicit farming of big cats for profit being exposed in Thailand (19), South Africa (20), and most recently the United States, which was popularized by the release of Netflix's *Tiger King*. In the United States, up until the recent (2022) passing of federal legislation (21; Big Cat Public Safety Act), there was very little oversight over the breeding, owning, and selling of tigers.

The pace of breeding in privately owned tiger populations has prompted concerns of inbreeding, especially since phenotypes (i.e., striped whites, snow whites, and golden tigers) that rely on rare, recessive mutations (22, 23) are considered to be of high economic value. In fact, the white tiger ("Bengal") breeding program was discontinued by the AZA when the intentional selection for the single variant responsible for producing the white phenotype (22, 23), was determined to be detrimental to conservation. The phenotype attracts unwanted attention from exotic breeders, and since it is a recessive mutation, the phenotype was generally achieved through inbreeding. This population was also intentionally mixed with Amur tigers in an attempt to prevent inbreeding depression before the program was ultimately discontinued (8). A previous study documented additional signatures of admixture in captive, privately owned tigers from various localities using mitochondrial DNA and microsatellites, and suggested that captive tigers contain unique diversity compared to their wild and captive single-ancestry counterparts (24). However, neither of the aforementioned studies on captive diversity (11, 24) examined the extent of inbreeding or genetic load in the population using genomic data, which has significantly more power than microsatellite data to answer such questions. It remains unclear how or if this large population of privately owned, captive tigers could fit into current management or conservation plans, whether their genomes hold relict or unique diversity, or if they show indications of severe inbreeding or high genetic load. As the largest source of tigers in the world, the captive, Generic population cannot be discounted as a source of genetic diversity and as an extended safeguard for tiger conservation (25). Individuals with single-subspecies ancestry also may exist in captive populations and the identification of these individuals may be critical to maintaining healthy gene pools of subspecies in captivity, should subspecies management remain as a priority in conservation (26). Mixed ancestry individuals from captive and semicaptive populations have now been used for the rewilding of wild carnivores including from mixed ancestry metapopulations of, for example, African wild dogs into Mozambique (27), cheetahs into Malawi (28), and for the carnivorous marsupial, the Tasmanian devil (29). Knowledge of the ancestry and diversity of the Generic tiger population is an essential and necessary step to truly consider them for roles in ex or in situ management decisions.

For the purposes of conservation genomics, the most useful definition of genetic load is one that describes a reduction in fitness at the population level due to deleterious mutations (30). The genetic load of a population can affect overall population health and viability, and can be modified by stochastic processes such as drift, admixture, and inbreeding (31, 32). Recent genomic-scale work in wild populations has shown purging of deleterious mutations in some groups (33–35), while other groups (36, 37) appear to have a large fraction of deleterious mutations remaining. In fact, recent work in wild Bengal tigers has shown purging of deleterious variation in smaller populations relative to larger populations, alongside continued inbreeding depression due to high frequency deleterious alleles (38). However, there is a dearth of literature on genetic load and inbreeding in captive tiger populations, or other wild tiger subspecies at the genomic scale. As genetic rescue, translocation, and captive breeding programs are being implemented with increasing frequency in endangered species management plans (39, 40), a broader understanding of genetic diversity and genetic load within captive and wild tiger populations warrants further investigation.

Here, we use genomic data to examine admixture and population structure, quantify genetic diversity and genetic load, and investigate the extent of inbreeding in the captive, privately owned tiger population in the United States (also known as "Generic" tigers) using whole-genome sequence data obtained from individuals in accredited sanctuaries. Tigers of unknown ancestry are considered Generic tigers and as having mixed ancestry unless genetic testing can confirm that they are from single ancestry sources. These sanctuaries where these samples were obtained do not breed, sell, or buy tigers, but instead house tigers previously rescued from unregulated facilities or those that have been forfeited by private owners and are ineligible to be incorporated into AZA breeding programs unless their ancestry can be verified. We combine our newly sequenced individuals with previously published data, resulting in a dataset representing 255 unique individuals across all six extant wild subspecies and U.S. Generic tigers (Table 1). These data allow us to determine how admixture events have shaped the genomic landscape of the privately owned captive population. We compare the Generic tigers to their potential source populations (tigers of single ancestry, i.e., subspecies), and examine how diversity is partitioned across each group. Then, we test for potential signs of inbreeding and quantify the total amount of deleterious variation in each population. Last, we show that ultralow coverage (0.25×) data can be imputed sufficiently using reference haplotypes from single ancestry tigers to determine ancestry and perform individual identification. Our results demonstrate that low-coverage sequencing and imputation provide a simple and cost-effective alternative compared to microsatellites

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

Table 1. Number of individuals (duplicates removed) used in each analysis by method (imputed/unimputed) with brief description of filters used

Group	Total individuals		Total	PCA	Admixture	Heterozygosity (unimputed, <20% missing)	ROH, IBD, Genetic Load (unrelated, >5×)	SFS (unrelated, >5×, even sampling)	Local ancestry (unimputed)
Amur	38	Unimputed	32	32	32	28	22	10 & 6	32
		Imputed	6	6	6				
Bengal	23	Unimputed	23	23	23	23	21	10 & 6	23
		Imputed	0	0	0				
Indochinese	6	Unimputed	6	6	6	6	6	6	6
		Imputed	0	0	0				
Malayan	22	Unimputed	22	22	22	22	20	10 & 6	22
		Imputed	0	0	0				
South China	4	Unimputed	1	1	1	1	1		
		Imputed	3	3	3				
Sumatran	18	Unimputed	18	18	18	18	15	10 & 6	18
		Imputed	0	0	0				
Generic	144	Unimputed	68	68	68	50	16	10 & 6	68
		Imputed	76	76	76				

See *Methods* and *SI Appendix, Supplementary Methods* for additional and specific details.

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

and custom single nucleotide polymorphism (SNP) panels to identify the source populations and identity of illegally traded individuals or wildlife materials from tigers and for long-term monitoring of captive and wild populations.

Results

Population Structure and Ancestry in Generic Tigers. Using both imputed and unimputed individuals, we investigated population structure and ancestry of the Generic tigers using a combination of principal component analysis (PCA) and supervised ADMIXTURE (see *SI Appendix, Supplementary Notes* for ancestry confirmation of single-subspecies individuals). Both the supervised ADMIXTURE analysis (Fig. 1*A* and *SI Appendix, Supplementary Notes*) and PCA (*SI Appendix*, Fig. S1), show that Generic tigers are admixed between the six extant tiger subspecies, with individuals varying in the proportion of their genome derived from each ancestral population (Fig. 1*A*). Investigation of mitochondrial haplotypes revealed clustering with all but the Sumatran and South China haplotypes (*SI Appendix*, Figs. S2–S4). Local ancestry analyses (unimputed samples only) revealed tracts from all subspecies that were analyzed (*SI Appendix*, Fig. S5). Overall, we found that none of the Generic tigers tested have single-subspecies ancestry. Generic individuals from the United States are primarily admixed between the Amur and Bengal tiger subspecies (mean: 38.6% and 28.2% of ancestry derived from each subspecies, respectively) and derive the least ancestry from the Sumatran and Malayan populations (mean: 11.1% and 5.0%, respectively; Fig. 1*A*). Our clustering analyses and PCA results showed that the captive population clusters together by their dominant ancestry in both PC space and clustering of identity-by-state measures (Fig. 1*B* and *SI Appendix*, Fig. S6*C*). We also observe some clustering based on putative geographic origin in the United States and Canada (Fig. 1*C*), though it is visually less distinguishable than the clustering by ancestry. These results suggest that structure of the Generic population is primarily caused by the breeding events that established the captive population in the United States and there is weak evidence for subsequent geographic structuring, suggesting that breeders are exchanging individuals. We see further evidence of historical admixture based on mitochondrial haplotype results, where Generic individuals cluster with Amur, Indochinese, and Bengal tiger subspecies. We observed only one Generic sample (SRR836354) grouping with Malayan tigers in mitochondrial haplotype analyses, and none grouping with Sumatran or South China tigers (*SI Appendix*, Figs. S2–S4). Further information can be found in *SI Appendix, Supplementary Notes*.

Interestingly, our analyses also identified several individuals previously labeled as single-ancestry subspecies to be Generic individuals. These individuals were originally labeled as Amur (SRR7651464-67, SRR7651470) and Bengal (SRR836354) subspecies in public repositories, but we observe that they all have a mixed ancestry background based on both PCA (*SI Appendix*, Fig. S17) and subsequent admixture analysis (Fig. 1*A*). These individuals lack additional metadata but are presumed to have originated in Asia (Dataset S1) and thus are labeled as "Asia" in the admixture plot (Fig. 1*A*).

Tiger Subspecies Diversity and Population History. Since the Generic tigers form a well-admixed population, we next investigated their genetic diversity, inbreeding, and mutation load relative to the single-ancestry, wild subspecies. Primarily, we wanted to determine whether Generic tigers were potential reservoirs of genetic diversity that is absent in the wild, as has been previously suggested. Given that bottlenecks have occurred in wild populations and during the process of founding the various captive tiger populations, the surviving diversity in the Generic population in comparison to wild populations is unknown. To compare genetic diversity across these groups, we estimated the number of distinct alleles per-locus, calculated the proportion of sites that are variant and segregating in the Generic tigers, but fixed in the wild subspecies, and calculated observed heterozygosity for each individual using only unimputed individuals (Fig. 2 and *SI Appendix*, Figs. S7 and S8). Allelic Diversity Analyzer (ADZE) (41) estimates allelic and shared diversity through a rarefaction approach. We found that trajectories for allelic diversity suggested the Indochinese subspecies (*SI Appendix*, Fig. S7) contained the most diversity, and were followed by the Bengal subspecies.

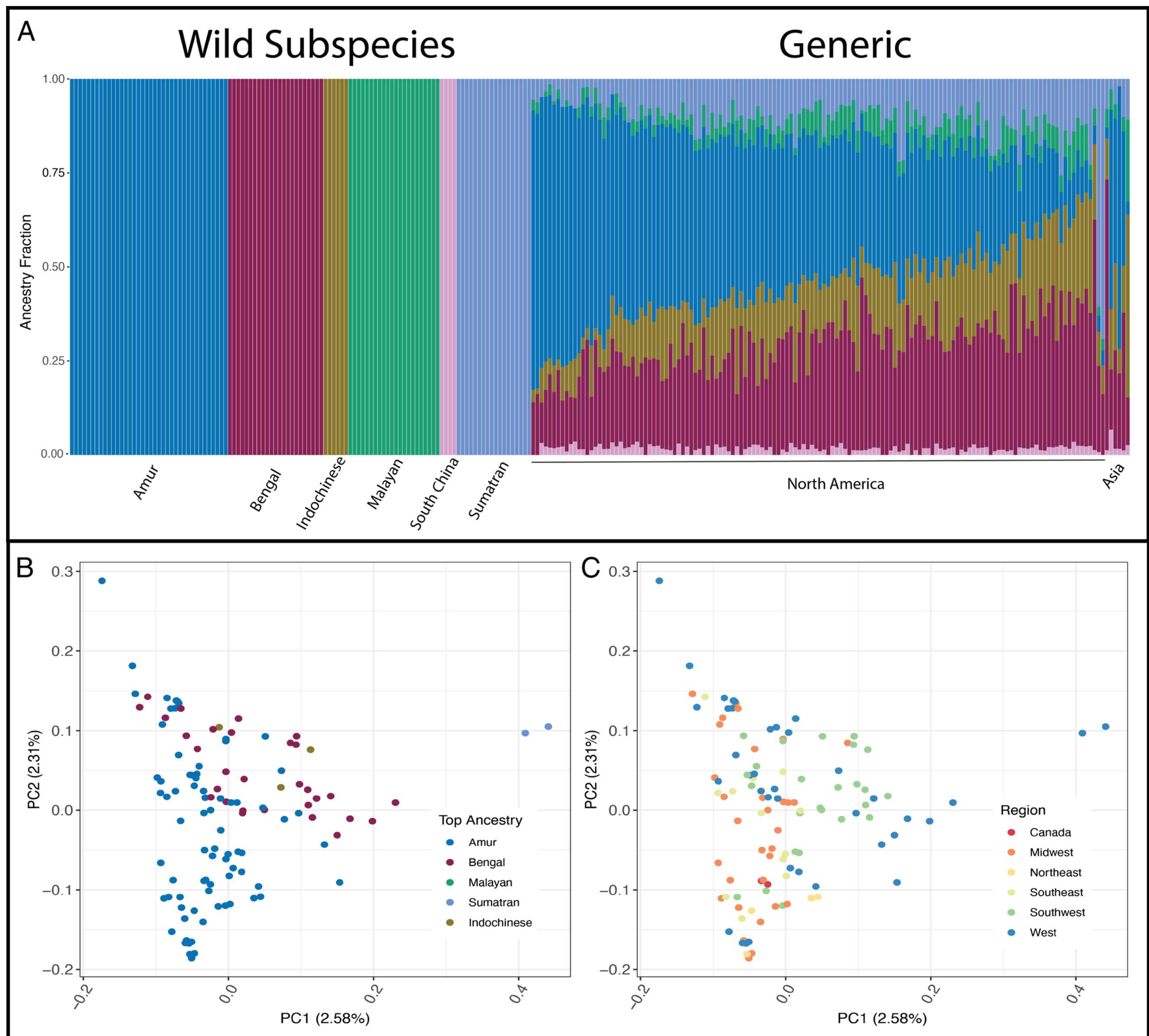

Fig. 1. Ancestral diversity in Generic tigers. (*A*) ADMIXTURE analysis showing ancestry of all individuals, wild and Generic. (*B*) PCA of North American Generic tigers where color corresponds to an individual's major ancestry component. (*C*) PCA of Generic tigers where color corresponds to the region of North America from which the individual originated. For panels *B* and *C* one individual was removed as an outlier; details can be found in *SI Appendix, Supplementary Notes*.

However, due to the small sample size for Indochinese tigers (N = 6), it is unclear whether their allelic diversity scales with sample size as it does in the Bengal tiger subspecies (*SI Appendix*, Fig. S8).

Next, we asked whether there might be an enrichment of variant sites that are solely segregating within the Generic population but fixed within any reference wild subspecies (Fig. 2*A*). We found that most sites that were segregating within the Generic tigers are also segregating in at least one of the wild subspecies (Fig. 2*A*). Specifically, less than 10% of sites are exclusively segregating within the Generic tigers (Fig. 2*A*). Finally, we examined heterozygosity. Generic tiger heterozygosity fell well within the range of other wild subspecies (Fig. 2*B*). Similar to what we observed with ADZE, Bengal and Indochinese tigers had the highest heterozygosity. Taken together, our analyses suggest that the Generic population is not a major reserve of unique genetic diversity, rather, it contains diversity that is present in wild subspecies of tigers.

Due to captive breeding practices and a small number of founders, we examined whether the Generic tiger population showed signs of inbreeding. Specifically, we estimated inbreeding using multiple metrics, including runs of homozygosity (ROH), shared identity-by-descent (IBD) segments, and inbreeding coefficients from both SNPs (F_{SNP}) and ROH (F_{ROH}) using only unimputed data. We observed that the Amur population had, on average, the most Type C (i.e., long runs, informative about recent inbreeding) ROH and highest IBD-sharing (Fig. 3 and *SI Appendix*, Figs. S9 and S10), suggesting that Amur tigers experienced more severe inbreeding than the Generic population. Further, Amur tigers also had the largest amount of Type B (i.e., intermediate length) ROH, which indicates a long-term, small population size (42). Indeed, wild Amur tigers were documented as having experienced an extreme bottleneck in the 1930s/1940s (11, 43), possibly explaining these extreme patterns compared to other subspecies.

When examining two other metrics of inbreeding (F_{SNP} and F_{ROH}) we observed that the Generic tiger population was once again not an outlier. Instead, most Generic individuals had similar inbreeding values to individuals in the other subspecies (Fig. 3). However, there are a few individuals within the Generic, Bengal, and Amur populations that are quite inbred. The varying levels

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

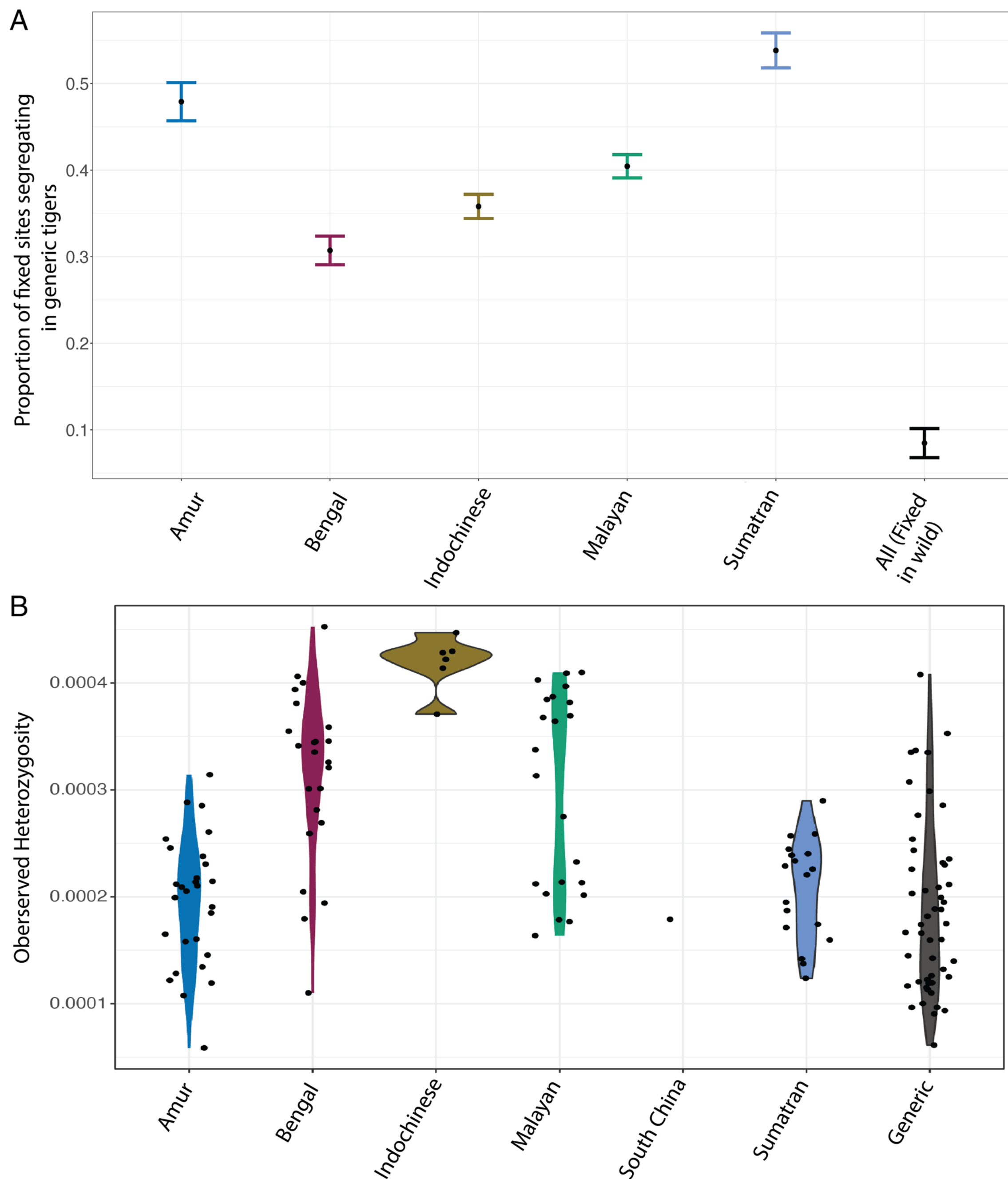

Fig. 2. Measurements of shared and within subspecies genetic diversity across tigers. (*A*) Proportion of segregating sites that are fixed in each wild subspecies compared to Generic tigers. The mean and SD are computed across 10 bootstrap samples of captive individuals (unimputed individuals). South China is not included since there is only one individual. The designation of "All" indicates the site is fixed across all wild subspecies but segregating in the Generic population. (*B*) Observed heterozygosity distributions across unimputed individuals, we have applied a horizontal jitter to each data point (See *SI Appendix, Supplementary Methods and Imputation and Filtering* for details).

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

of inbreeding mirrored the previous analysis of heterozygosity, where we observed variable levels of genome-wide heterozygosity across the Generic population (Fig. 2*B*). The large variance in inbreeding coefficients (Type C ROH, F_{ROH}, and F_{SNP}) suggests that in the case of Generic tigers, admixture between distinct subspecies likely increased heterozygosity in some individuals, but inbreeding within individual facilities eroded it (44). Taken together, our results show that the Generic population is not any more inbred than wild tiger populations, nor is it more diverse.

Tiger Subspecies Mutation Load. We last sought to characterize the prevalence of putatively deleterious variation in the captive population, relative to each subspecies (unimputed samples). Variants were categorized as putatively deleterious if they were nonsynonymous (NS), derived mutations with a Sorting Intolerant from Tolerant (SIFT) score less than 0.05 (see *SI Appendix, Supplementary Methods* for more detail). We implemented multiple approaches to count deleterious variants in the genome: 1) counting homozygous derived genotypes; 2) counting the total number of homozygous and heterozygous derived genotypes (counting variants); and 3) counting alleles where we tabulate twice the number of homozygous derived genotypes plus heterozygous genotypes (counting alleles).

Given that the Generic population primarily derived its ancestry from the Amur and Bengal subspecies (Fig. 1*A*), we expected that the distribution of deleterious variation in Generic tigers

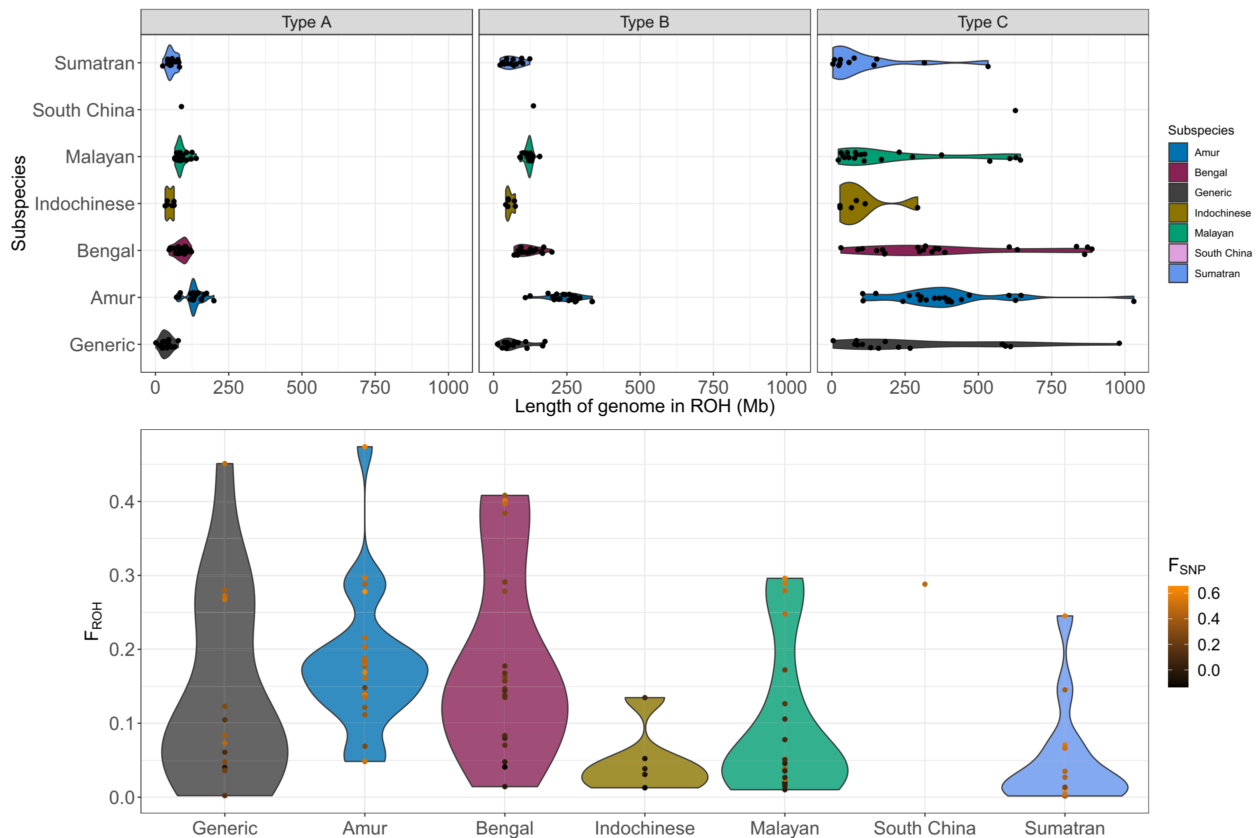

Fig. 3. Quantification of different types of ROH across tiger subspecies and measurements of inbreeding. (*Top*) Length of different classes of ROH (A; short, B; intermediate, C; long) for each tiger subspecies and Generic tigers. (*Bottom*) Total proportion of genome in Type C ROH (F_{ROH}) for each individual per population. We have labeled each individual with their corresponding inbreeding coefficient measured as F_{SNP}, the darkest color (black) is the lowest value, and the brightest color (orange) corresponds to the largest value.

would be comparable to the Bengal and Amur subspecies. Indeed, the observed distribution of both neutral and deleterious variants in Generic tigers was more similar to the Amur tigers and Bengal tigers than other wild subspecies (Fig. 4 and *SI Appendix*, Fig. S12). Overall, we observed similar levels of deleterious variation across Sumatran, Malayan, Indochinese, and Bengal subspecies. Though the mean number of derived deleterious variants was slightly, but consistently, higher in the Sumatran subspecies for all counting methods (*SI Appendix*, Fig. S12). We found that, irrespective of subspecies, individuals with the most putatively deleterious derived homozygotes also tended to have the largest inbreeding coefficients, quantified with either F_{SNP} or F_{ROH} (*SI Appendix*, Fig. S14). Thus, inbreeding likely led to an overall decrease in fitness for all tigers, irrespective of their ancestry. Our observation that the Generic tigers had the lowest average numbers of derived deleterious mutations suggests that the Generic tiger population has purged some of the variants that are deleterious in homozygous form being carried by the other subspecies, or the admixture process reduced their frequency as homozygotes.

Next, we examined whether putatively deleterious and nonsynonymous (NS) homozygous variation was enriched within ROH versus outside of ROH. Recall that counting homozygotes is relevant when we are interested in variants that are recessive. Here, we are asking whether there is an enrichment of putatively deleterious and NS within ROH, which would result in unmasked recessive variation and simultaneously expose the variant to selection. We identified a significant enrichment of both NS and putatively deleterious derived homozygotes within ROH versus outside of ROH across all subspecies (*SI Appendix*, Fig. S15). We saw that the Generic tigers were a clear standout in terms of carrying NS derived homozygotes within ROH (odds ratio ~2.9) relative to outside of ROH. However, the odds of putatively deleterious derived homozygotes falling within ROH in the Generic tigers are comparable to any of the wild subspecies. In the case of putatively deleterious derived homozygotes within ROH, the standout became the Indochinese subspecies, though it also had the largest 95% CIs (*SI Appendix*, Fig. S15).

Creation of a Reference Panel for Future Tiger Conservation. Given that we collated the largest genomic dataset of tigers to date, we also tested whether the data were sufficient to create an accurate reference panel for imputation of low-coverage samples. A reference panel can be particularly useful in wildlife forensics where the ancestry or identity of the individual is unknown, but also for cost-effective sequencing of many individuals at lower coverages when paired with imputation. Imputation is generally used in the context of low-coverage data to increase the reliability of genotype calls and improve the accuracy of downstream analyses, and has recently been advocated as an alternative to reduced representation (e.g., Restriction Site Associated DNA Sequencing [RADseq] and its derivatives) approaches in nonmodel species, which often amplify a small and potentially biased portion of the genome (45). The primary goal of the reference panel was to assess whether we could accurately impute ultralow coverage data for the purpose of individual identification and ancestry determination in tigers. The full data was split into a training set (reference panel) composed of 106 single-ancestry individuals (5, 6) with coverage between 2× and 43×. Then, imputation was performed using *lowimpute* on 86 low- and ultralow coverage (0.25× to 6×) samples (*SI Appendix*, Fig. S1). We tested accuracy in a set of nine individuals, where three of the

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

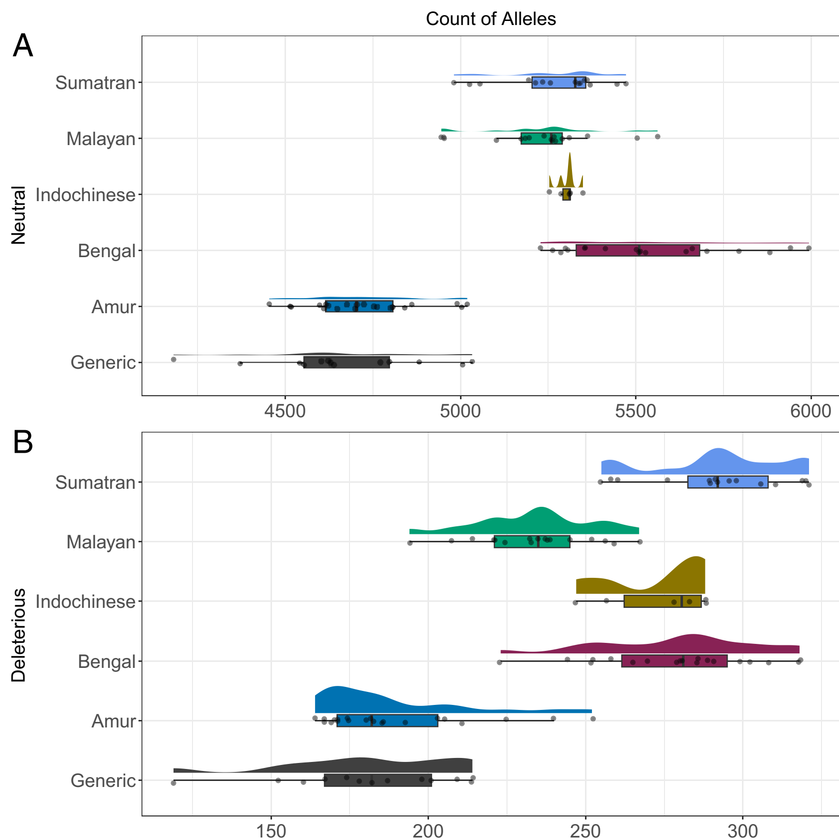

Fig. 4. Quantifying genetic load in each subspecies by counting derived neutral and deleterious variants. (*A*): Scaled counts of derived neutral alleles in each tiger subspecies and Generic tigers. (*B*): Scaled counts of derived deleterious alleles in each tiger subspecies and Generic tigers. Extended figure (*SI Appendix*, Fig. S12) contains the count of homozygotes and count of variants.

individuals were also represented in the reference (See *SI Appendix, Supplementary Notes* for additional details). To test these samples, we randomly downsampled reads from each individual to five different coverages five times (0.25×, 0.5×, 1×, 2×) in addition to no downsampling (using all the reads in a standard variant calling pipeline without imputation). Downsampled reads were run-through the imputation pipeline, and accuracy of the imputation was tested using variant calls (via nonreference discordance; NDR), ancestry predictions, and the ability to uniquely identify individuals.

We found that as sequencing depth increased, the accuracy of variant calls, as measured by NDR, also increased (Dataset S2). Average NDR was 16.48% at the lowest depth (0.25×) and 9.06% at the highest depth (2×) (*SI Appendix*, Fig. S16 and Dataset S3). In general, samples with lower overall coverage had higher NDR rates, which can be attributed to less overall variant call reliability at lower depths (*SI Appendix*, Fig. S16 and Dataset S3). Despite the variance in NDR across samples (max NDR at 0.25× was 28.79% in one sample), ancestry (*SI Appendix*, Fig. S17) and the ability to identify individuals (Dataset S4), even at ultralow coverage (0.25×), remained accurate and we found that there were no significant differences in ancestry assignments across the replicates at various depths (Dataset S4). Our results show that low-coverage sequencing and imputation are a sufficient and low-cost alternative to high-coverage sequencing or genotyping for the purposes of ancestry and individual identification using this reference panel for tigers.

Discussion

Admixture and Variation. We found that all Generic tigers sampled from the United States are admixed and contain ancestry from all six extant tiger subspecies (Fig. 1*A*). Most Generic tigers derive a majority of their ancestry from the Amur and Bengal subspecies (Fig. 1*A*), but individuals varied in the proportion of ancestry derived from any one subspecies (Fig. 1*A* and *SI Appendix*, Fig. S5). The length distribution of local ancestry segments (*SI Appendix*, Fig. S5) indicated that there are likely very few individuals in the extant U.S. Generic tiger population that have directly descended from recently introduced, single-ancestry sources, and there appears to be little geographic structuring across the United States (Fig. 1*C*). Interestingly, a study examining privately owned individuals of sable antelope (*Hippotragus niger*) compared to zoo individuals found that zoo individuals had significantly higher genetic diversity and lower levels of inbreeding than individuals in privately owned populations (46). The study also found that zoo and privately held sable populations formed distinct clusters, likely due to genetic drift (46). This is despite the fact that, similar to the Generic tiger population, the number of sable antelope in privately owned facilities far exceeds the number in zoos. Though we did not have the opportunity to include sequencing of the extant populations of tigers in zoos, it will be critical to do so in the future to understand how and whether drift has affected these populations similarly and whether zoo populations have retained more diversity than the generic population. However, we find this unlikely, given that some of the single subspecies reference individuals in our dataset are the founders of the U.S. zoo populations (6).

While there is no doubt that the Generic tiger population carries some unique variation because of rapid growth and the large number of individuals in captivity (i.e., census size), we have shown in our sampling, that this population does not contain more variation or unique genetic variation compared to single-ancestry populations of tigers (Fig. 2 and *SI Appendix*,

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

Figs. S7 and S8). Our result is contrary to previous findings (24). Our work shows that Generic tigers have average heterozygosity (Fig. 2*B*), possess very few unique segregating sites (*SI Appendix*, Figs. S7 and S8), and when compared to other subspecies, contain sites that are also segregating in the wild populations (Fig. 2*A*). While there may be some advantageous alleles circulating in the Generic population, predicting which alleles are adaptive is a challenging task for even the most well-studied species, and implementing gene-based management strategies can also bring about harmful side-effects (47, 48).

Deleterious Variation. The Generic population contains less deleterious recessive homozygous variation and relatively low amounts of Class C ROH relative to wild populations (Figs. 3 and 4). Given the degree of admixture in the Generics, the low frequency of deleterious mutations is expected, yet surprising because of the purported inbreeding of "farmed" tigers (49). However, Generic tigers also have the largest enrichment of NS variation within ROH versus outside. In other words, though Generic tigers have a comparable amount of NS and putatively derived homozygotes compared to their wild ancestors, they tend to carry more of the NS variation within ROH than their ancestors. Thus, the results of admixture on the genomic landscape in Generic tigers are, in this case, not obvious. Heterozygosity and genetic load are similar across all tigers, likely due to a composite effect of admixture between distinct lineages increasing heterozygosity, followed by severe bottlenecks of repeated founding events and deliberate inbreeding. The observed outcome of admixture in the Generics is the result of chance but is not necessarily predictive of other metrics of inbreeding, load, or heterozygosity. The increase in NS load within versus outside of ROH, and population-level heterozygosity falling between major ancestry sources has also been observed in human populations, specifically in certain admixed Latin American founder populations (44).

Ancestry Inference for Unknown Individuals. The final objective of this study was to compile a reference dataset and imputation panel capable of identifying the ancestry and identity of individuals sequenced at ultralow coverages. At present, tiger forensic work relies primarily on mitochondrial and microsatellite data (50). Mitochondrial markers in felid lineages contain nuclear mitochondrial inserts (numts), making interrogating these regions difficult and error-prone (51, 52). Further, mitochondrial markers are only indicative of the maternal lineage of an individual and are not able to identify individuals with mixed ancestry. For complex admixture scenarios, examining only a handful of microsatellites may cause biased results (since microsatellites in nonmodel systems are often selected randomly or based on microsatellites in other species) or in the case of species or subspecies-level hybridization distort ancestry signals (53–56).

Notably, our findings using whole-genome data differ from the results obtained from (24), where 30 microsatellite loci and mitochondrial (4 kb) data were used to investigate samples from captive tigers (N = 105) obtained between 1982 and 2002 from 14 countries (including the United States). By comparing their data to voucher specimens (39 Amur; 2 South China; 33 Indochinese; 22 Malayan; 17 Bengal; 21 Sumatran), they found that 49 of the captive tigers belonged to single subspecies ancestry, while 52 were found to be of admixed ancestry (24). A total of 18 of 105 tigers did not match the suspected ancestry (i.e., the tiger was found to be admixed compared to the source reporting single subspecies ancestry or vice versa). Luo et al. (24) did not describe the origin of the sampled individuals in detail, but it is possible that individuals of single-subspecies ancestry were being actively mixed into the privately owned captive tiger population when those samples were collected. Alternatively, it is possible that the small number of microsatellites used in the previous study have comparatively limited power to distinguish the complex ancestry of the Generic tigers compared to thousands of SNPs, as has been the case in other systems (41–43). Interestingly, Luo et al. (24) found that captive tigers with mixed subspecies ancestry primarily contained ancestry from the Indochinese tiger lineage, whereas the tigers analyzed herein contained mostly Amur and Bengal ancestry, which could possibly be attributed to samples being obtained from captive populations outside of the United States.

The reference set of individuals described here is notably smaller than those used in other imputation studies from model organisms due to the difficulty in obtaining wild tiger specimens, though studies have successfully implemented imputation using only 44 reference individuals (57). However, because F_{ST} values between tiger subspecies are generally high (~0.2; 5, 6), we show that the panel can be used accurately for analyzing ancestry and identifying individuals, which are two of the most common objectives for wildlife forensics and conservation management work. We do not test this panel for haplotype-level analyses such as inferring ROH or precise analyses such as genetic load, since these analyses generally require large reference panels or higher coverage genomic sequencing. Further, we note that the reference panel can and should be improved overtime, especially for populations where sample sizes are low (e.g., South China and Indochinese tigers) and populations which are known to contain population structure not sampled here (e.g., Bengal tigers; 38).

Conclusions

There has been ongoing discussion regarding how captive populations might contribute to conservation, and particularly whether they harbor unique alleles beneficial to the survival of the species or represent historical variation that is no longer present in the wild. Here, we have shown that the captive, privately owned tiger population does not contain significant unique variation compared to the wild subspecies. Additional investigations into the remaining diversity in captive U.S. zoo populations are unlikely to reveal additional diversity since many of these lineages are the putative founders of the Generic tiger population. Several notable zoo populations, such as giraffes (58) and orangutans (59) have sourced animals for captive breeding programs from multiple, genetically distinct populations, subspecies, or species. If we continue to consider captive populations as diversity reservoirs for wild species, additional investigation into the potential outcomes and interplay of such processes (admixture, inbreeding versus outbreeding depression) is extremely important.

Cumulatively, our analyses are concordant with the known history of tiger subspecies in terms of historic bottlenecks and recent inbreeding and provide a comprehensive picture of the diversity across all extant subspecies using available data. None of the Generic tigers had single subspecies ancestry, indicating a history of breeding practices in captive tigers inconsistent with that of AZA policy. Indeed, most Generic tigers contain ancestry from all six wild tiger subspecies in their genomes. Contrary to previous hypotheses, most of the studied Generic tigers do not show signs of severe, recent inbreeding, nor do they hold unique diversity. Thus, the role they might play (if any) for tiger conservation is unclear. We find no obvious genetic reasons why Generic tigers from the United States would not be suitable in augmenting tiger populations, since most individuals do not contain an excess of deleterious mutations or appear to be inbred. However, genetic results are only a subset of the considerations made during genetic rescue and population augmentation decisions (60). Whether it

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

is wise to keep the tiger subspecies separate to preserve their genetic uniqueness, or whether certain circumstances warrant lineage mixing remains to be seen for tigers and other species as populations decline and diversity is eroded. South China tigers have already been genetically rescued by another subspecies (61), and what populations may require for survival is more individuals, but where they are sourced is a critical choice. For example, here we showed that the Sumatran tiger population in particular may be in need of conservation action due to an excess of homozygous deleterious alleles (Fig. 4 and *SI Appendix*, Figs. S12 and S13). Simulations may help determine the best sources for possible translocations (38). However, more work must be done to validate analyses of deleterious variation and improve our understanding of the genotype to fitness map in tigers before suggestions are integrated into conservation planning, as these analyses can be imprecise even in the most well-studied organisms (62–64).

To encourage future study and aid illegal trade and trafficking investigations, we also present a reference panel that in conjunction with imputation can accurately identify the ancestry and identity of an individual using even ultralow coverage sequencing, demonstrating the validity of this approach. As sequencing costs decline, arrays and panels are set to emerge as the predominant and cost-effective way to query forensic and low-quality samples, compared to microsatellites (52, 53). Arrays and panels require either a sufficient number of SNPs (54, 55) or if a reduced set is needed to be cost-effective, a very careful selection of SNPs which are reflective of ancestry or relatedness. Low-coverage WGS is an attractive alternative to creating marker panels, which often are affordable and accessible only with caveats (e.g., when running many markers or many samples), that often exclude conservation budgets and sample needs. Overall, our results are pertinent to other globally threatened species which are held in captivity, showing specifically that genomic studies can help resolve long-held misconceptions (such as those about population diversity and inbreeding) and provide information pertinent to future conservation efforts.

Methods

Sample Collection and Sequencing. A total of 154 tiger samples were collected opportunistically during routine vet care from sanctuary facilities by vet and sanctuary staff or from existing biobank collections (Dataset S1). All samples were extracted using a Qiagen DNeasy kit (Cat. No. 69504) and samples prepared using a modified Nextera library prep protocol (65). 77 of these samples (listed as "Unimputed" in Dataset S1) were sequenced between approximately 2× and 5× depth. The remaining 77 samples ("Imputed" in Dataset S1) were sequenced at approximately 0.25×. Tigers collected in sanctuaries or with unknown ancestry are labeled as "Generic" for the purposes of this manuscript. We additionally collected data on the putative location of birth of each tiger from the sanctuaries in North America. These locations were then translated into one of six regions: West (Washington, OR, CA, NV, MT, ID, WY, CO, UT), Southwest (AZ, NM, TX, OK), Midwest (ND, SD, NE, KS, MN, IA, MO, WI, IL, MI, IN, OH), Southeast (AR, LA, TN, MS, AL, GA, FL, KY, WV, VA, NC, SC), and Northeast (ME, NH, MA, NY, PA, MD, DE, CT, VT, RI, NJ), and Canada (Canada was not subdivided since N = 2 and both individuals were from the same facility).

Variant-Calling and Reference Panel Construction. An additional 100, publicly available (as of December 2019) tiger genome samples were downloaded from NCBI. Reads were mapped to the GenTig1.0 genome (66) using BWA-MEM v0.7.17 (67) and variant calling was subsequently performed by Gencove using the Genome Analysis Toolkit v4.1.4.1 (68) according to best practices (69). Initial variant calling was performed on all samples available at the time, excluding those sequenced at 0.25×. All of these samples are referred to as unimputed for the purposes of this manuscript. Initial variant calling yielded a total of 23,579,569 variable sites across the entire genome. We restricted calls to biallelic sites using BCFtools v1.6 (70), and subsequently filtered for quality, missingness, and depth. See *SI Appendix, Supplementary Methods* for additional details. The final dataset contained a total of 7,519,430 sites across unimputed individuals and a callable genome of 2,174,711,735 base pairs.

In order to select individuals to build the reference panel and accurately split individuals into groups for kinship estimation, we conducted PCA to ensure that all individuals in the unimputed dataset were clustering according to subspecies using PLINK v2 (71). Individuals were subsequently split into ancestry groups to form the reference panel, which included representatives from all six tiger subspecies, but no Generic individuals. Further, we tested several methods for detecting relatedness using pedigreed individuals in the dataset, which were subsequently used to identify and remove duplicates. Additional information can be found in *SI Appendix, Supplementary Methods*.

Imputation and Filtering. Using only single-subspecies ancestry individuals verified here and by previous studies, we developed a reference panel to impute variants for an additional 86 individuals (labeled as imputed in Dataset S1) through the *loimpute* pipeline developed by Gencove and available at www.gencove.com (72). The 86 imputed individuals were composed primarily of individuals sequenced at ultralow coverage (N = 75; 0.25×), but also included two individuals from a Canadian Zoo sequenced at approximately 3×, and an additional nine samples that became publicly available after the initial variant calling had been performed (see Dataset S1 for details).

We combined the variant call files (VCFs) for imputed individuals with the unimputed individuals using BCFtools *merge*. Because imputation emits a call for every site in the reference pipeline, we restricted the merged sites VCF to retain only the quality sites identified after initial variant calling and filtering. We further checked for imputation accuracy using concordance measures, specifically NDR (73), and examined the accuracy of ancestry and relatedness measures over a variety of coverages. Additional information can be found in *SI Appendix, Supplementary Methods*.

Genetic Diversity. We used several approaches to investigate genetic diversity across tigers. The first approach was to create equal-sized groups of (N = 10) individuals and tabulate the proportion of sites that were SNPs in the Generic population and fixed in each wild subspecies. We used the same reference groups from each subspecies and generated 10 replicate samples (with replacement) of the Generic tigers to check whether these proportions varied across the Generic population.

Observed heterozygosity was also calculated as the total number of heterozygous sites divided by the number of callable sites (2,174,711,735) in the genome. Observed homozygous sites were counted in each subspecies using VCFtools (74) using the "--het" flag and heterozygous sites were calculated by subtracting the (O)HOM (observed homozygosity) column from the NSITES (the total sites queried) column. The number of callable sites was determined as the total number of base pairs minus the sites with mappability scores <1 for autosomal scaffolds. We additionally tested to see whether heterozygosity was correlated with missingness. See *SI Appendix, Supplementary Methods* for details.

ROH. GARLIC (75) was used to detect ROH. The error was set at 0.001, the window size at 700, and centromeres were set as 0,0 since no centromere information was available. To ensure that ROH were not mistakenly called on regions with an excess of missing calls, each ROH was then intersected with the number of callable sites in each window (see *SI Appendix, Heterozygosity* section for details). ROH larger than 100 kb and containing callable sites within one SD (±0.066) of the mean coverage (0.655) were retained. ROH were binned into different size classes A (short), B (intermediate), and C (long) per subspecies. Binning was based on the use of a Gaussian mixture function that fits a model to the ROH length distribution within the group. Type A ROH are typically indicative of linkage-disequilibrium blocks. Type B ROH are informative about long-term small population sizes and cryptic relatedness. The presence of Type C ROH indicates recent inbreeding in the population. F_{ROH} was computed as the total fraction of the genome within a type C ROH.

IBD Segments. IBD segments were called using TRUFFLEv1.38 with parameters "--segments --missing 1 --maf 0 --nofiltering" (76). Segments greater than 2 Mb where the fraction of coverage by callable sites (count listed above) was within one SD (±0.032) of the mean coverage (0.660) were retained. IBD scores were

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

computed using the same approach from Nakatsuka et al. (77), where we normalize shared IBD by the number of sampled individuals.

Putatively Neutral and Deleterious Variation. We polarized and annotated sites by first filtering the data to scaffolds that corresponded to main chromosomes from felCat8 (78; GCF_000181335.2). Coordinates were identified using liftOver (79). The remaining sites, with felCat8 coordinates, were annotated with an impact and consequence using Ensembl Variant Effect Predictor (VEP) v92. We removed intergenic sites, splice acceptor, splice donor, splice region annotations, and selected the most damaging impact for a given transcript. Sites were coded as NS, synonymous (SYN), or loss of function (LOF) and SIFT scores were added (80). Information from VEP was combined with a SIFT score to find putatively neutral (SYN with SIFT score greater than 0.05) and putatively deleterious sites (NS or LOF with SIFT score less than 0.05). See *SI Appendix, Supplementary Methods* for details about annotations and SIFT scores.

Ancestral bases were identified with Progressive CACTUS (81) alignment of the GenTig1.0 genome with nine other felid genomes; further details can be found in *SI Appendix, Supplementary Methods*. To assess load in each subspecies, we used only unimputed individuals with at least 5× coverage and less than 5% missing data. Last, we scaled the number of sites per individual by subtracting the total number of variant sites across all individuals from missing sites to get the total number of called sites. Each count was divided by the number of callable sites for that individual; the proportion was multiplied by the average number of callable sites across all subspecies.

Data, Materials, and Software Availability. All code associated with this project can be found at https://doi.org/10.5281/zenodo.13540809 (82). Whole genome sequence data associated with this study has been deposited into NCBI under bioproject number PRJNA976043 (83). Raw VCFs have been uploaded to dryad repository https://doi.org/10.5061/dryad.k0p2ngff1 (84).

ACKNOWLEDGMENTS. We would like to acknowledge each of the sanctuaries and zoos who contributed samples, knowledge, and time to this study, specifically the Performing Animal Welfare Society (PAWS), the Exotic Feline Rescue Center (EFRC), In-Sync Exotics, Turpentine Creek Wildlife Refuge, Carolina Tiger, Oakland Zoo, and others. We especially acknowledge Joe Taft and Rebecca Rizzo of EFRC, Jackie Gai and Ed Stewart of PAWS, and Vicky Keahy of In-Sync, who have spent many hours discussing this project with us. We thank Tara Harris of the AZA for assistance and knowledge regarding the SSP programs and sample access. This project would not have been possible without Bill Nimmo and Kizmin Reeves of Tigers in America. In addition, we thank the San Diego Frozen Zoo for access to their collections and their collaborative spirit. We also thank Sara Huston Katsanis for procuring the Carolina Tiger Rescue and Tiger World specimens and supporting the mentoring grant for V. Grant.

Author affiliations: [a]Department of Biology, Stanford University, Stanford, CA 94305; [b]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089; [c]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806; [d]Department of Genetics, School of Medine, Stanford University, Stanford, CA 94305; [e]Department of Ecology, Evolution & Behavior, The Hebrew University of Jerusalem, Jerusalem 9190500, Israel; [f]Gencove, Inc., New York, NY 11101; [g]San Diego Zoo Wildlife Alliance, Escondido, CA 92027; [h]National Centre for Biological Sciences, Tata Institute for Fundamental Research, Bangalore 560065, India; [i]Chan Zuckerberg BioHub, San Francisco, CA 94158; [j]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305; [k]Department of Earth System Science, Stanford University, Stanford, CA 94305; [l]Woods Institute for the Environment, Stanford University, Stanford, CA 94305; and [m]Center for Innovation in Global Health, Stanford University, Stanford, CA 94305

Author contributions: E.E.A., U.R., D.A.P., and E.A.H. designed research; E.E.A., J.A.M., K.P., and T.Y. performed research; G.S.B., V.B.G., G.G., C.B.K., O.A.R., and T.Y. contributed new reagents/analytic tools; E.E.A., J.A.M., K.A.S., B.Y.K., G.G., K.P., and J.K.P. analyzed data; and E.E.A., J.A.M., K.A.S., B.Y.K., J.K.P., N.R., D.A.P., and E.A.H. wrote the paper.

1. R. Dirzo *et al.*, Defaunation in the anthropocene. *Science* **345**, 401–406 (2014).
2. J. Walston *et al.*, Bringing the Tiger back from the brink–The six percent solution. *PLoS Biol.* **8**, e1000485 (2010).
3. R. Tilson, K. Traylor-Holzer, Q. M. Jiang, The decline and impending extinction of the South China tiger. *Oryx* **31**, 243–252 (1997).
4. H.-R. Xue *et al.*, Genetic ancestry of the extinct Javan and Bali tigers. *J. Hered.* **106**, 247–257 (2015).
5. Y.-C. Liu *et al.*, Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Curr. Biol.* **28**, 3840–3849 (2018).
6. E. E. Armstrong *et al.*, Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. *Mol. Biol. Evol.* **38**, 2366–2379 (2021).
7. S.-J. Luo, Y.-C. Liu, X. Xu, Tigers of the World: Genomics and conservation. *Annu. Rev. Anim. Biosci.* **7**, 521–548 (2019).
8. R. Tilson, P. J. Nyhus, *Tigers of the World: The Science, Politics and Conservation of Panthera tigris* (Academic Press, 2009).
9. D. G. Kleiman, K. V. Thompson, C. K. Baer, *Wild Mammals in Captivity: Principles and Techniques for Zoo Management* (University of Chicago Press, ed. 2, 2010).
10. K. Traylor-Holzer, *AZA North American Regional Amur Tiger Studbook* (Minnesota Zoo/IUCN Conservation Breeding Specialist, 2007).
11. P. Henry *et al.*, In situ population structure and ex situ representation of the endangered Amur tiger. *Mol. Ecol.* **18**, 3173–3184 (2009).
12. K. Ralls, J. D. Ballou, "Managing genetic diversity in captive breeding and reintroduction programs" in *Transactions of the 57th North American Wildlife and Natural Resources Conference* (Wildlife Management Institute, Washington, D.C., 1992), pp. 263–282.
13. D. J. Fraser, How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evol. Appl.* **1**, 535–586 (2008).
14. J. R. Willoughby *et al.*, The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. *Mol. Ecol.* **24**, 98–110 (2015).
15. H. Ito, R. Ogden, T. Langenhorst, M. Inoue-Murayama, Contrasting results from molecular and pedigree-based population diversity measures in captive zebra highlight challenges facing genetic management of zoo populations. *Zoo. Biol.* **36**, 87–94 (2017).
16. R. Ogden *et al.*, Benefits and pitfalls of captive conservation genetic management: Evaluating diversity in scimitar-horned oryx to support reintroduction planning. *Biol. Conserv.* **241**, 108244 (2020).
17. B. R. Wright, C. J. Hogg, E. A. McLennan, K. Belov, C. E. Grueber, Assessing evolutionary processes over time in a conservation breeding program: A combined approach using molecular data, simulations and pedigree analysis. *Biodivers. Conserv.* **30**, 1011–1029 (2021).
18. R. Garner, A census of federally regulated big cat populations within the United States as of december 2020. *J. Zoolog. Botan. Gard.* **2**, 517–530 (2021).
19. E. Cohen, The fall and reincarnation of Thailand's Tiger Temple. *Int. J. Tourism Anthropol.* **7**, 115–131 (2019).
20. V. L. Williams, A. J. Loveridge, D. J. Newton, D. W. Macdonald, A roaring trade? The legal trade in Panthera leo bones from Africa to East-Southeast Asia. *PLoS One* **12**, e0185996 (2017).
21. GovInfo, "Big cat public safety act" (Public Law 117-243, US Government Publishing Office, 2023; https://www.govinfo.gov/content/pkg/COMPS-17151/pdf/COMPS-17151.pdf).
22. X. Xu *et al.*, The genetic basis of white tigers. *Curr. Biol.* **23**, 1031–1035 (2013).
23. X. Xu *et al.*, The genetics of tiger pelage color variations. *Cell Res.* **27**, 954–957 (2017).
24. S.-J. Luo *et al.*, Subspecies genetic assignments of worldwide captive tigers increase conservation value of captive populations. *Curr. Biol.* **18**, 592–596 (2008).
25. R. Tilson, U. Seal, *Tigers of the World: The Biology, Biopolitics, Management, and Conservation of an Endangered Species* (William Andrew Publishing, 1987).
26. S. Maity *et al.*, DNA matchmaking in captive facilities: A case study with tigers. *Mol. Biol. Rep.* **49**, 4107–4114 (2022).
27. P. Bouley, A. Paulo, M. Angela, C. Du Plessis, D. G. Marneweck, The successful reintroduction of African wild dogs (Lycaon pictus) to Gorongosa National Park, Mozambique. *PLoS One* **16**, e0249860 (2021).
28. O. Sievert, J. Fattebert, K. Marnewick, A. Leslie, Assessing the success of the first cheetah reintroduction in Malawi. *Oryx* **56**, 505–513 (2022).
29. E. A. McLennan, C. E. Grueber, P. Wise, K. Belov, C. J. Hogg, Mixing genetically differentiated populations successfully boosts diversity of an endangered carnivore. *Anim. Conserv.* **23**, 700–712 (2020).
30. G. Bertorelle *et al.*, Genetic load: Genomic estimates and applications in non-model animals. *Nat. Rev. Genet.* **23**, 492–503 (2022).
31. D. Charlesworth, J. H. Willis, The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
32. L. F. Keller, D. M. Waller, Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241 (2002).
33. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in Island foxes. *Curr. Biol.* **28**, 3487–3494.e4 (2018).
34. C. Grossen, F. Guillaume, L. F. Keller, D. Croll, Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nat. Commun.* **11**, 1001 (2020).
35. Y. Xue *et al.*, Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).
36. L. Smeds, H. Ellegren, From high masked to high realized genetic load in inbred Scandinavian wolves. *Mol. Ecol.* **32**, 1567–1580 (2023).
37. J. A. Mooney, C. D. Marsden, A. Yohannes, R. K. Wayne, K. E. Lohmueller, Long-term Small Population Size, Deleterious Variation, and Altitude Adaptation in the Ethiopian Wolf, a Severely Endangered Canid. *Mol. Biol. Evol.* **40**, msac277 (2023).
38. A. Khan *et al.*, Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2023018118 (2021).
39. A. R. Whiteley, S. W. Fitzpatrick, W. C. Funk, D. A. Tallmon, Genetic rescue to the rescue. *Trends Ecol. Evol.* **30**, 42–49 (2015).
40. K. A. Witzenberger, A. Hochkirch, Ex situ conservation genetics: A review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. *Biodivers. Conserv.* **20**, 1843–1861 (2011).
41. Z. A. Szpiech, M. Jakobsson, N. A. Rosenberg, ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498–2504 (2008).
42. T. J. Pemberton *et al.*, Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
43. S. Alasaad, R. C. Soriguer, G. Chelomina, Y. P. Sushitsky, J. Fickel, Siberian tiger's recent population bottleneck in the Russian Far East revealed by microsatellite markers. *Mamm. Biol.* **76**, 722–726 (2011).
44. J. A. Mooney *et al.*, Understanding the hidden complexity of latin American population isolates. *Am. J. Hum. Genet.* **103**, 707–726 (2018).
45. M. M. Watowich *et al.*, Best practices for genotype imputation from low-coverage sequencing data in natural populations. *Mol. Ecol. Resour.* 1–13 (2023), 10.1111/1755-0998.13854.

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.

46. R. M. Gooley *et al.*, Comparison of genomic diversity and structure of sable antelope (Hippotragus niger) in zoos, conservation centers, and private ranches in North America. *Evol. Appl.* **13**, 2143-2154 (2020).
47. M. Kardos, A. B. A. Shafer, The peril of gene-targeted conservation. *Trends Ecol. Evol.* **33**, 827–839 (2018).
48. P. S. Miller, P. W. Hedrick, MHC polymorphism and the design of captive breeding programs: Simple solutions are not the answer. *Conserv. Biol.* **5**, 556–558 (1991).
49. J. Guo, Tigers in trouble: Year of the tiger. *Nature* **449**, 16–18 (2007).
50. D. Karmacharya *et al.*, Species, sex and geo-location identification of seized tiger (Panthera tigris tigris) parts in Nepal—A molecular forensic approach. *PLoS One* **13**, e0201639 (2018).
51. J.-H. Kim *et al.*, Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene* **366**, 292–302 (2006).
52. W. Zhang *et al.*, Highly conserved D-loop-like nuclear mitochondrial sequences (Numts) in tiger (Panthera tigris). *J. Genet.* **85**, 107–116 (2006).
53. L. Szatmári *et al.*, A comparison of microsatellites and genome-wide SNPs for the detection of admixture brings the first molecular evidence for hybridization between Mustela eversmanii and M. putorius (Mustelidae, Carnivora). *Evol. Appl.* **14**, 2286–2304 (2021).
54. A. Miralles *et al.*, Inconsistent estimates of hybridization frequency in newts revealed by SNPs and microsatellites. *Conserv. Genet.* **25**, 215–225 (2024).
55. L. K. Vaughan *et al.*, The use of plasmodes as a supplement to simulations: A simple example evaluating individual admixture estimation methodologies. *Comput. Stat. Data Anal.* **53**, 1755–1766 (2009).
56. B. M. Fitzpatrick, Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evol. Biol.* **12**, 131 (2012).
57. Z. L. Fuller *et al.*, Population genetics of the coral Acropora millepora: Toward genomic prediction of bleaching. *Science* **369**, eaba4674 (2020).
58. L. B. Lackey, *Giraffe Studbook Giraffa Camelopardalis North American Regional/Global* (Association of Zoos and Aquariums and Disney's, 2011).
59. G. L. Banes, B. M. F. Galdikas, L. Vigilant, Reintroduction of confiscated and displaced mammals risks outbreeding and introgression in natural populations, as evidenced by orang-utans of divergent subspecies. *Sci. Rep.* **6**, 22026 (2016).
60. M. F. Breed *et al.*, The potential of genomics for restoring ecosystems and biodiversity. *Nat. Rev. Genet.* **20**, 615–628 (2019).
61. W. Zhang *et al.*, Sorting out the genetic background of the last surviving South China tigers. *J. Hered.* **110**, 641–650 (2019).
62. M. Kayser, P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.* **12**, 179–192 (2011).
63. N. A. Rosenberg, L. M. Li, R. Ward, J. K. Pritchard, Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* **73**, 1402–1422 (2003).
64. N. A. Rosenberg, Algorithms for selecting informative marker panels for population assignment. *J. Comput. Biol.* **12**, 1183–1201 (2005).
65. M. Baym *et al.*, Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**, e0128036 (2015).
66. E. E. Armstrong *et al.*, Genome Report: Chromosome-level draft assemblies of the snow leopard, African leopard, and tiger (Panthera uncia, Panthera pardus pardus, and Panthera tigris). *G3 (Bethesda)* **12**, 6762081 (2022).
67. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arxiv [Preprint] (2013). https://arxiv.org/abs/1303.3997 (Accessed 1 January 2018).
68. A. McKenna *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
69. H. Li *et al.*, Exome variant discrepancies due to reference-genome differences. *Am. J. Hum. Genet.* **108**, 1239–1250 (2021).
70. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
71. C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
72. K. Wasik *et al.*, Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics* **22**, 197 (2021).
73. R. Arthur, J. O'Connell, O. Schulz-Trieglaff, A. J. Cox, Rapid genotype refinement for whole-genome sequencing data using multi-variate normal distributions. *Bioinformatics* **32**, 2306–2312 (2016).
74. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
75. Z. A. Szpiech, A. Blant, T. J. Pemberton, GARLIC: Genomic autozygosity regions likelihood-based inference and classification. *Bioinformatics* **33**, 2059–2062 (2017).
76. A. Dimitromanolakis, A. D. Paterson, L. Sun, Fast and accurate shared segment detection and relatedness estimation in un-phased genetic data via TRUFFLE. *Am. J. Hum. Genet.* **105**, 78–88 (2019).
77. N. Nakatsuka *et al.*, The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* **49**, 1403–1407 (2017).
78. J. U. Pontius *et al.*, Initial sequence and comparative analysis of the cat genome. *Genome Res.* **17**, 1675–1689 (2007).
79. M. Haeussler *et al.*, The UCSC Genome browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
80. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
81. J. Armstrong *et al.*, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
82. J. Mooney, E. Armstrong, jaam92/Tigers: Tigers (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.13540809. Deposited 29 August 2024.
83. E. Armstrong *et al.*, Data from "Unraveling the genomic diversity and admixture history of captive tigers in the United States." National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA976043/. Deposited 24 May 2023.
84. E. Armstrong *et al.*, Data from: Unraveling the genomic diversity and admixture history of captive tigers in the United States [Dataset]. Dryad. https://doi.org/10.5061/dryad.k0p2ngff1. Deposited 10 October 2023.

Downloaded from https://www.pnas.org by STANFORD UNIVERSITY on September 20, 2024 from IP address 171.66.13.154.