Divergent reference genomes compromise the reconstruction of demographic histories, selection scans, and population genetic summary statistics

Maria Akopyan[1], Matthew Genchev[1], Ellie E. Armstrong[2], and Jazlyn A. Mooney[1]

1. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA
2. Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, Riverside, CA, USA

Correspondence: J.A.M., jazlynmo@usc.edu

Abstract

Characterizing genetic variation in natural populations is central to evolutionary biology. However, most non-model organisms lack integral genomic resources such as reference genomes and recombination maps, limiting accurate evolutionary inference. Here, we explore the consequences of reference genome bias on the inference of genetic diversity, demographic histories, and recombination rates using gray foxes (*Urocyon cinereoargenteus*), which, like most members of *Canidae,* are traditionally mapped to the dog (*Canis lupus familiaris*) reference genome. Whole genome sequence data from gray foxes were mapped to the gray fox reference genome and two heterospecific canid references (dog and Arctic fox; *Vulpes lagopus*). Our results reveal that reference bias significantly affects population genomic analyses. Mapping to the conspecific gray fox genome improved read pairing, increased detection of SNPs, especially rare variants, and reduced spurious variants. Estimates of nucleotide diversity (π) and genetic differentiation (F_{ST}) were higher using the gray fox genome. We observed that mapping to heterospecific references leads to underestimates of population sizes, distorted demographic trajectories, and more variable recombination rates. These effects are further complicated by population-specific biases, which vary in their magnitude and direction across populations, highlighting the need for tailored approaches to mitigate reference bias. Importantly, F_{ST} outlier detection also differed among references, affecting functional interpretations. Collectively, this work addresses a critical gap in the rapidly expanding field of non-model species genomics by demonstrating the importance of using conspecific genomic resources in evolutionary research and illustrating how reliance on distantly related reference genomes can distort population genetic analyses.

Introduction

Accurately characterizing genetic variation is essential for understanding the evolutionary dynamics of populations and informing management strategies for species of conservation concern. The increasing accessibility of whole genome sequencing (WGS) data for non-model organisms has led to a proliferation of studies employing WGS data to estimate genetic diversity, assess population structure and connectivity, and identify adaptive genetic variation. Recent advancements in analytical techniques for WGS data have enabled sophisticated analyses that extend beyond classic population genetic diversity metrics to characterize structural variation[1], infer demographic histories[2], and estimate population recombination rates[3]. These developments provide deeper insights into the historical and contemporary processes shaping genetic variation in natural populations. However, the reliability of these inferences and

cross-population comparisons depends on the reference genome used. Mapping reads to a divergent reference genome can skew our understanding of genetic variation, because a single reference genome cannot adequately capture the breadth of genetic diversity within or between species.

Reference bias can occur when a single reference genome–typically a haploid sequence from a single individual–is used to represent the diversity of an entire species, sometimes leading to errors in mapping or variant calling. For example, in ancient human data, reference bias has resulted in misestimation of heterozygosity[4], shared genetic drift[5], allele frequency, and ancestry[6]. Reference bias has also been shown to affect genetic diversity estimators, distributions of commonly used neutrality statistics, and divergence in gorillas[7]. In some cases, conspecific reference genomes are unavailable, and researchers turn to reference genomes from related species. Similar results for misestimated divergence were observed in modern dogs and wolves[8]. In non-model species, like big cats, both heterozygosity and demographic histories changed when a conspecific reference was used[9]. In fishes, reference bias has led to overestimated heterozygosity[10], inaccurate detection of structural variation[11], and distorted phylogenies[12], potentially skewing conclusions about population history and genetic diversity. Lastly, recent work on the impact of reference bias in two species of conservation concern, the New Zealand kiwi and the beluga whale, revealed alarming mismatches in demography and inhibited the detection of long runs of homozygosity[13].

Despite the impact of reference bias on population genetic inference, few studies have examined how it may affect the characterization of demographic history and population recombination rates, which are increasingly popular targets of estimation. The gray fox (*Urocyon cinereoargenteus*) provides a compelling case study in this research. Gray foxes are widely distributed across North America, and despite their phenotypic similarity, genetic evidence suggests deep divergence between eastern and western lineages[14]. Previous gray fox genetic studies have relied on a domestic dog (*Canis lupus familiaris*) reference genome[15], which poses challenges due to significant karyotypic differences between the two species[16,17]. In fact, despite most canid study's reliance on the domestic dog reference genome, the clade contains a number of large rearrangements, for example, dogs have 38 pairs of autosomal chromosomes, gray foxes have 32, and the Arctic fox (*Vulpes lagopus*) has 24. These differences in synteny and chromosome structure are likely to bias downstream haplotype-based analyses that rely on synteny and linkage disequilibrium (e.g., demographic inference and recombination mapping). In addition to large-scale karyotypic changes across *Canidae*, this family is home to both common (gray fox), near-threatened (Channel Island fox)[18,19], and endangered (Ethiopian wolf)[20] species. Despite these genomic differences, most previous genomic research within *Canidae* has been conducted by mapping to the domestic dog, primarily because it was the only canid reference genome available with annotations and had both high continuity and contiguity[21].

The recently published gray fox reference genome[17] provides us with a high-quality conspecific reference for alignment, variant calling, and ultimately quantifying reference bias. Here, we re-analyze previously published WGS data from two populations of North American gray foxes[15]. We examine how inference of demographic histories and recombination landscapes differ when using the heterospecific domestic dog Arctic fox genomes in contrast to the conspecific gray fox genome. To investigate the effects of reference bias on our analysis, we compare the underlying genetic variation and the site frequency spectrum (SFS) estimated

for each population using different reference genomes, as well as estimates of genetic diversity and differentiation. Finally, we examine the effect of reference bias on both coalescent and SFS based demographic inference as well as F_{ST} outlier scans for selection and recombination rate inference.

Results

We analyzed whole-genome resequencing data from 12 gray foxes, including six from the eastern U.S. and six from the western U.S. populations (Figure S1 and 1A). We examined reference bias by mapping these data to three reference genomes: the domestic dog genome CanFam4[22], the Arctic fox genome[23], and the conspecific gray fox genome[17] (Figure 1B).

Read pairing is improved with a conspecific reference genome

We found significant differences in the proportion of reads mapped between the conspecific and heterospecific reference genomes. Approximately 99% of gray fox reads successfully mapped to the Arctic fox and CanFam4 reference genomes, while 99.7% of gray fox reads mapped to the gray fox reference genome (χ^2_{KW}=33.19, P<0.001). We found no significant differences in the estimated depth across individuals between the genomes, resulting in average autosomal read depths of 9.34× for the Arctic fox, 8.78× when using Canfam4, and 8.66× when using the gray fox reference (χ^2_{KW}=0.98, P=0.613). Interestingly, we found a substantial and significant difference in the percentage of properly paired reads across genomes, with 89.4%, 90.3%, and 94.7% of reads properly pairing when mapped to the Arctic fox, Canfam4, and the gray fox reference genome, respectively (χ^2_{KW}=40.14, P<0.001). On average, almost 5% more reads were properly paired when mapped to the gray fox reference compared to the two heterospecific genomes.

A species-matched reference genome yields more SNPs and rare variants

We examined the impact of reference genome choice on the detection and characterization of genetic variation and allele frequency distributions in gray fox populations. Across all reference genomes, we detected more SNPs in the western population of gray foxes compared to the eastern population, with nearly double the number of SNPs identified in the west (Figure S2). Notably, the gray fox reference genome consistently yielded in the calling of high-quality SNPs compared to the heterospecific references, with 13,199,882 variants identified in the west, compared to 10,012,122 and 9,969,686 variants detected in the same western individuals with the Arctic fox and dog genomes, respectively. The species-matched reference yielded more SNPs in the east as well, with 6,828,177 variants, compared to 5,407,943 and 5,392,973 variants identified based on the Arctic fox and dog genomes, respectively.

We identified more singletons (i.e., rare variants where an allele is only found once in the population) in the east and west using the gray fox reference (Figure 1C). In addition, mean SNP depth varied slightly across genomes, with gray fox showing higher values (mean: east; 8.5, west; 8.7) compared to Arctic fox and dog (both east: 8.3, west: 8.6). Missing data on a per-

site basis was lowest with the gray fox genome (east and west: 0.02), while the Arctic fox and dog genomes had higher missingness (east: 0.05, west: 0.03).

After lifting over SNPs in the heterospecific genomes to their corresponding positions in the gray fox genome, we found that approximately 80% of the variants identified using heterospecific reference genomes were also identified as variants when using the gray fox genome (Figure S2). Among these matching variants, non-singleton SNPs made up about 50-53%, while the remainder were singletons. The west population had a higher proportion of matching singletons compared to the east (~32% vs. ~24%), whereas the east population had a slightly higher proportion of matching non-singleton SNPs (~53% vs. ~50%). For the variants that did not lift over as SNPs, the majority (14-18%) were not identified as any type of variant in the gray fox genome.

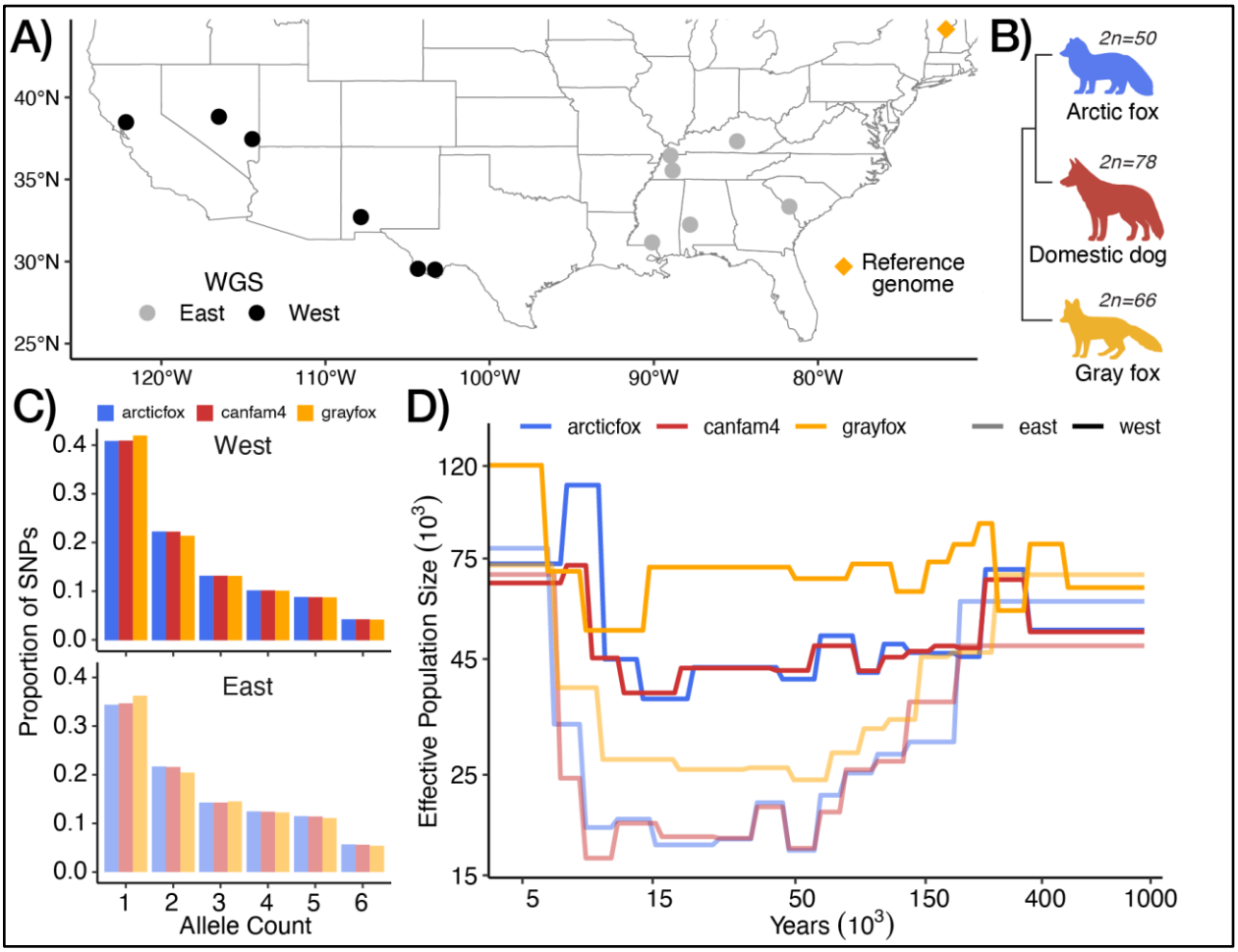

Figure 1. Reference bias influences the SFS and demographic trajectories. A) *Sampling localities of gray fox genomes, with a star indicating the origin of the gray fox reference genome[17] and circles representing WGS data obtained from Preckler-Quisquater et al.[15].* ***B)*** *Silhouettes and phylogenetic relationships of the three canid species whose reference genomes were used: Arctic fox (Vulpes lagopus) in blue, domestic dog (Canis lupus familiaris) in red, and the most basal clade, gray fox (Urocyon cinereoargenteus) in gold. Each species is labeled with its respective diploid chromosome number (2n).* ***C)*** *SFS for eastern and western populations show more singletons using the conspecific gray fox genome.* ***D)*** *Inferred effective population sizes (y-axis) and years from present (x-axis) reveal discordant demographic histories of eastern (light) and western (dark) foxes resolved using the species-matched (gold) and heterospecific genomes (red and blue).*

A species-matched reference genome results in higher effective population size estimates

We inferred demographic histories using smc++ and observed distinct population size trajectories when using different reference genomes for both the eastern and western populations. Across all genomes, the western population consistently showed higher effective population size (N_e) estimates than the eastern population. The gray fox genome produced trajectories with smaller fluctuations in N_e over time and generally higher N_e estimates, particularly for the western population. In contrast, the Arctic fox and dog genomes revealed more variability, especially in the west, where the discrepancies between reference genomes were more pronounced. For instance, in the most recent time period, approximately 5,000-7,000 years ago, the gray fox genome indicated population growth in the west, with N_e increasing from 70,000 to 120,000. Meanwhile, the heterospecific genomes suggested a population decline, with N_e estimates dropping below 75,000.

To evaluate the extent to which different demographic inference methods are impacted by reference bias, we supplemented the smc++ analysis, which combines the SFS with linkage disequilibrium (LD) information in coalescent hidden Markov models (HMMs), with MSMC2, based only on a coalescent HMM, and Stairway Plot2, which relies solely on the SFS (Figure S3). Both MSMC2 and Stairway Plot 2 exhibited greater inconsistencies across reference genomes for the eastern population compared to the western population. With MSMC2, the eastern population showed recent growth across all references, but this growth was highly exaggerated with the heterospecific references, resulting in population size estimates for the east surpassing those for the west—an outcome not observed with the other methods and inconsistent with diversity estimates, which are higher in the west (discussed below). In contrast, Stairway Plot 2 revealed larger discrepancies in the eastern population during ancient times. For the western population, MSMC2 inferred recent growth only when using heterospecific references, whereas Stairway Plot 2 indicated stability across references, highlighting the varying sensitivities of demographic inference methods to reference bias.

Recombination rates vary across reference genomes and populations

We detected significant differences in the distributions of recombination rates inferred using the different reference genomes in both populations (Anderson-Darling k-sample test, $p<0.001$). In the eastern population, recombination rates were lower when inferred using the Arctic fox genome compared to the gray fox genome. The average recombination rate was 0.448 centiMorgans per megabase (cM/Mb) with the Arctic fox genome, which is approximately 31% lower than the 0.650 cM/Mb observed with the gray fox genome. Conversely, when using the CanFam4 genome, the average recombination rate increased to 0.872 cM/Mb, about 34% higher than the rate inferred with the gray fox genome. In the western population, recombination rates were higher when inferred using both heterospecific genomes compared to the conspecific gray fox genome. Specifically, the average recombination rate was 1.02 cM/Mb with the Arctic fox genome, approximately 13% higher than the 0.903 cM/Mb observed with the gray fox genome, and 0.989 cM/Mb with the CanFam4 genome, about 9.5% higher than the rate inferred

with the gray fox genome. Overall, these differences were more pronounced in the eastern population, where the heterospecific genomes led to both underestimation (Arctic fox genome) and overestimation (CanFam4 genome) of recombination rates relative to the gray fox genome. In contrast, in the west, both heterospecific genomes consistently overestimated recombination rates compared to the gray fox genome, and the differences were less pronounced.

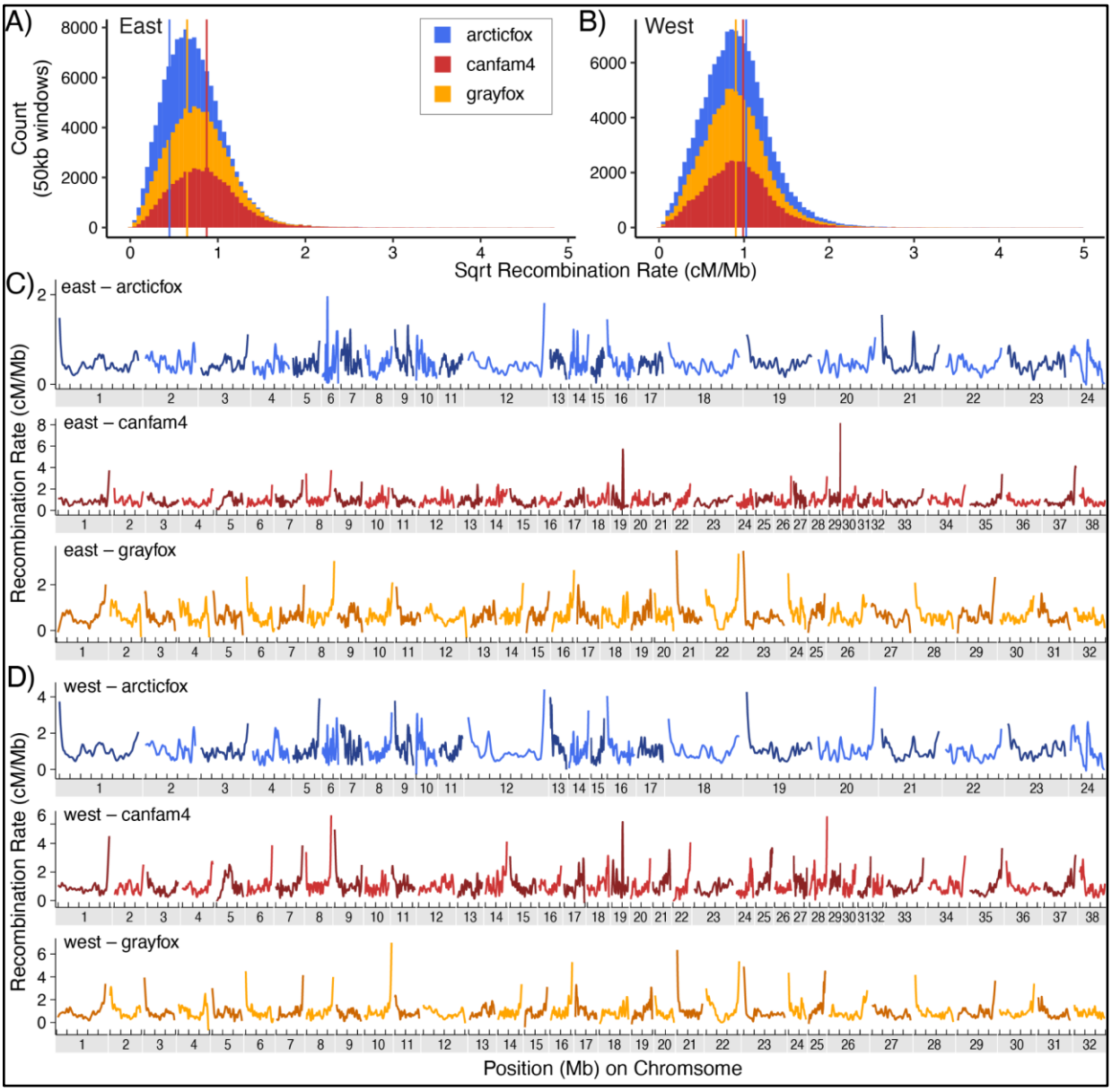

Figure 2. Recombination rate comparisons between reference genomes for gray fox populations. *Histograms of the square root of recombination rates (cM/Mb) across 50 kb windows in (A) east and (B) west populations. Colors represent the three reference genomes used for recombination rate estimates, with vertical lines representing mean values. Loess-smoothed (span=0.1) recombination rates (cm/Mb) per chromosome computed over 50 kb windows in Eastern (C) and Western (D) gray foxes based on the three reference genomes.*

Additionally, recombination landscapes based on each genome revealed substantial variation across 50 kb windows, with heterospecific reference genomes resulting in increased variability, particularly toward chromosome ends. In the eastern population, using the conspecific gray fox genome, recombination rates ranged from 0.000343 to 7.34 cM/Mb. However, when using the heterospecific Arctic fox genome, the maximum recombination rate doubled to 14.9 cM/Mb, and with the CanFam4 genome, it more than tripled to 23.4 cM/Mb. Similarly, in the western population, recombination rates estimated with the gray fox genome varied from 0.000598 to 11.6 cM/Mb. When the Arctic fox genome was used, the maximum recombination rate more than doubled to 24.9 cM/Mb, and with the CanFam4 genome, it increased to 14.5 cM/Mb. Compared to the gray fox genome, the heterospecific genomes resulted in higher maximum recombination rates, with a consistent trend of overestimation toward the ends of chromosomes, suggesting that heterospecific reference genomes can lead to an overestimation of recombination rates, especially at the higher end of the spectrum.

Heterospecific genomes underestimate diversity and differentiation

We detected significant differences in estimates of average nucleotide diversity (π) in 50 kb windows in both gray fox populations based on the three reference genomes (Figure 3). In the eastern population, nucleotide diversity was lowest with the CanFam4 genome (π=0.000618), followed closely by Arctic fox (π=0.000625), while the gray fox genome yielded a higher estimate of π=0.000812 (χ^2_{KW} = 6663.3, P<0.001). Similarly, in the western population, diversity was lowest using the heterospecific references (Arctic fox: π=0.00122, CanFam4: π=0.00121), while the gray fox genome produced a higher estimate of π=0.00164 (χ^2_{KW}=12409, P<0.001).

The variation in estimates of nucleotide diversity was slightly more evident in the west population, where π was approximately 1.34 times higher using the gray fox genome compared to the heterospecific references, while in the east, π was 1.31 times higher. While estimates of π were higher in the west than in the east across all reference genomes, the difference between populations was greater when using the gray fox genome, where diversity in the west was about 2.02 times higher than in the east. By contrast, diversity based on the Arctic fox genome showed a 1.95-fold difference between the west and east, and CanFam4 showed a similar 1.96-fold difference.

Estimates of genetic differentiation (F_{ST}) between the eastern and western populations also varied depending on the reference genome. The heterospecific references produced identical F_{ST} values (mean F_{ST}=0.189), while the gray fox genome resulted in a significantly higher mean F_{ST} of 0.197 (χ^2_{KW}=558.27, P<0.001). Additionally, the correlation between nucleotide diversity and F_{ST} was more strongly negative in the eastern population (τ=-0.173 to -0.196) compared to the western population (τ=-0.081 to -0.118), with the strongest correlations observed using the gray fox reference in both populations.

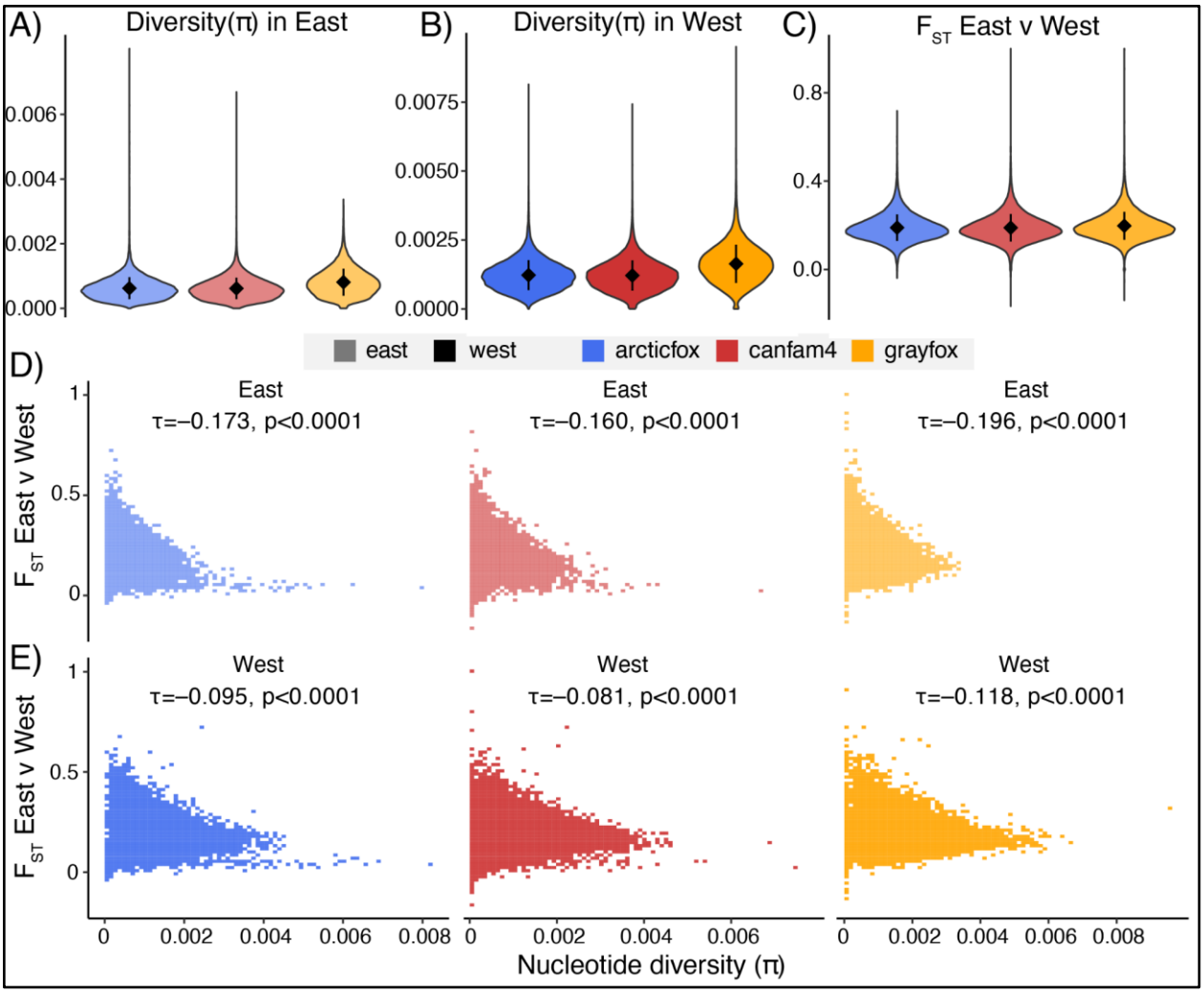

Figure 3. Estimates of diversity and differentiation are higher with gray fox reference in eastern and western gray fox populations. *Nucleotide diversity (π) for eastern (A) and western (B) gray fox populations, and genetic differentiation (F_{ST}) between the populations (C), in 50-kb windows based on three reference genomes. Violin plots with mean estimates and standard deviation error bars are shown. The correlation between diversity and differentiation across genomic windows for the east (D) and west (E) populations is illustrated. Kendall's tau (τ) correlation coefficients are reported, showing significant negative correlations between π and F_{ST} in both populations across all reference genomes.*

Reference genome choice affects F_{ST} outlier detection

We defined outliers as windows with F_{ST} values exceeding three standard deviations above the mean for each reference genome, then matched these windows across reference genomes to identify shared and unique outlier regions. Across comparisons, we observed that heterospecific references consistently identified more than twice the number of unique outlier windows compared to the conspecific reference (Figure 4). Notably, the number of windows unique to each heterospecific reference was similar to the number of shared outliers between the heterospecific and conspecific references. In the Arctic fox and gray fox comparison, 148 shared outlier windows were identified, with 137 unique to the Arctic fox reference and 63

unique to the gray fox reference (Figure 4A). Between the Canfam4 and gray fox reference genomes, 165 outlier windows were shared, with 141 unique to Canfam4 and 61 unique to gray fox (Figure 4B). The majority of windows in both comparisons did not show elevated F_{ST} values in either reference (33,350 and 31,649 in the Canfam4 and the Arctic fox comparisons, respectively), reflecting our conservative method of outlier detection.

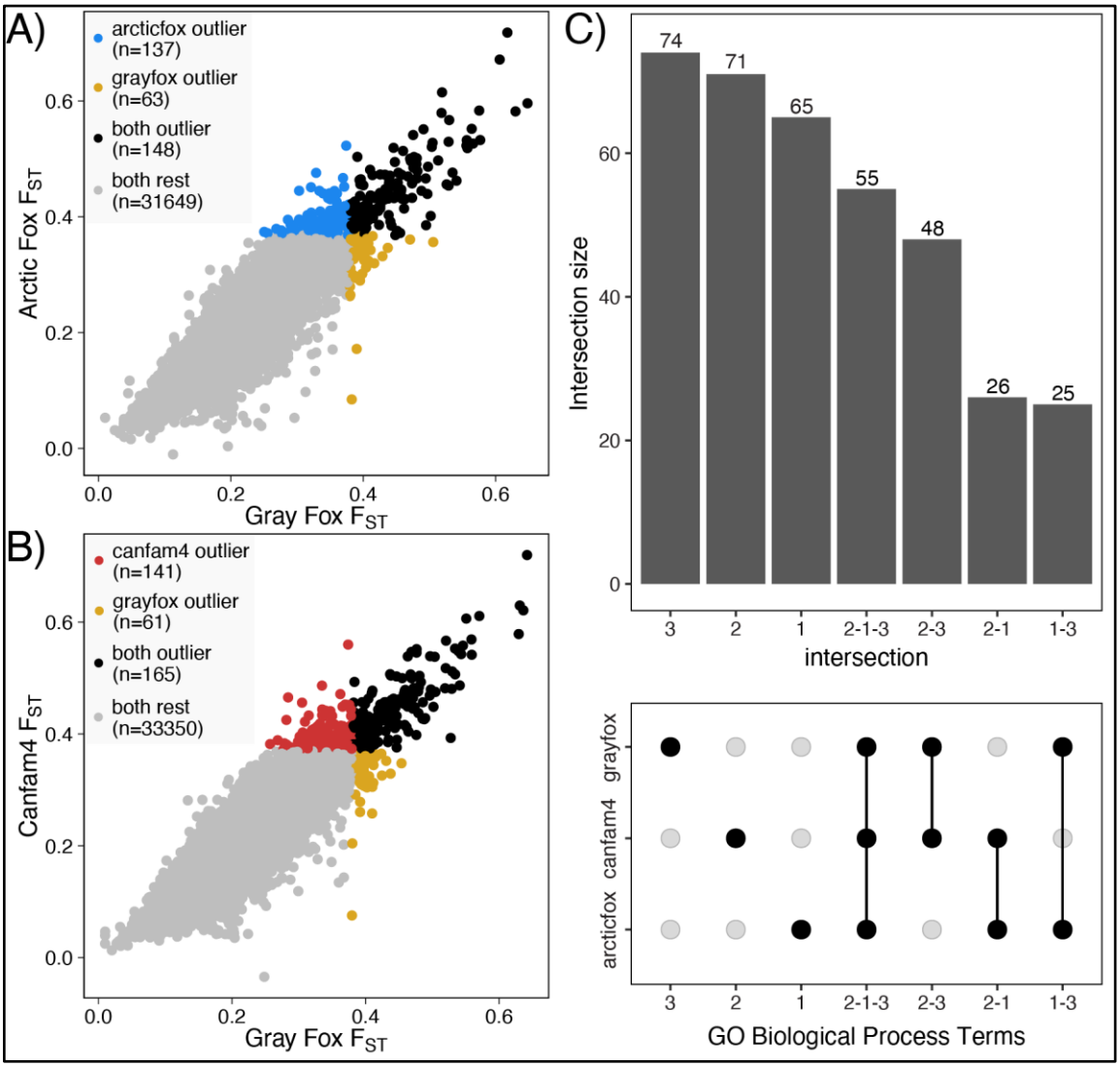

Figure 4. Reference-specific patterns revealed by F_{ST} outlier and functional enrichment analyses. *(A) Scatter plot comparing F_{ST} values between gray fox and arctic fox references, with outlier windows unique to the arctic fox (blue) and gray fox (gold) highlighted. (B) Scatter plot comparing F_{ST} values between gray fox and Canfam4 references, highlighting outliers unique to Canfam4 (red) and gray fox (gold). For both (A) and (B), outliers shared between the two references are shown in black, while non-outlier windows are in gray. (C) UpSet plot of significant (p<0.05) GO biological process terms identified as enriched for each reference genome. Bars represent the intersection size of GO terms. Shared terms across all three references are represented by the bar labeled "2-1-3," while pairwise intersections show shared terms between specific reference genomes. The bottom panel shows the intersections represented as dots, where each connected set of dots indicates a shared subset of GO terms across references.*

Functional enrichment patterns reveal reference-specific biological processes in F_{ST} outliers

The functional enrichment analysis of F_{ST} outliers revealed distinct and overlapping gene ontology (GO) biological process terms across reference genomes (Figure 4C). Gray fox had the highest number of unique terms with 74, followed by Canfam4 with 71, and Arctic fox with 65, indicating that each reference captured distinct biological processes. Importantly, there were more unique terms than shared terms across references. Among shared terms, 55 were common to all three references. Notably, the overlap between gray fox and Canfam4 (48 shared terms) was larger than the overlap between Arctic fox and either gray fox (25 terms) or Canfam4 (26 terms), suggesting a closer alignment in identified biological processes between the gray fox and Canfam4 references.

Further, the unique terms identified by each reference genome suggested distinct thematic focuses (Figure S4). The Arctic fox reference highlighted processes related to cellular transport, secretion, and signaling, with terms like "positive regulation of catecholamine secretion" and "golgi to endosome transport," indicating an emphasis on cellular signaling and developmental regulation. The Canfam4 reference captured growth, differentiation, and endocrine response, with terms such as "positive regulation of muscle cell differentiation" and "response to thyroid hormone," pointing to processes involved in physiological development. In contrast, the gray fox reference emphasized nervous system function, immune modulation, and cell proliferation, reflected by terms like "positive regulation of nervous system process" and "calcineurin-mediated signaling."

Discussion

Reference genomes are essential tools in genetic studies, serving as coordinate systems for annotating sequence features and comparing individuals and populations. However, many non-model organisms lack annotated, high-quality chromosome-level reference genomes[24]. In the absence of conspecific reference genomes, researchers often use genomes from closely related species[25,26], which can compound reference bias. Reference bias has been previously studied across the tree of life[8,9,13,27–30] both within species[28,29] and between species[9]. In this study, we demonstrate that mapping to a heterospecific reference genome significantly impacts population genomic inferences. This leads to discrepancies in effective population size estimates, recombination rates, measures of genetic diversity, differentiation, and selection signals. Our findings highlight the importance of using conspecific genomes for accurate evolutionary inferences.

Although average sequencing depth showed only minor discrepancies using a discordant reference, we observed significant differences in both the percentage of reads mapped and the proportion of properly paired reads. Reads are not properly paired when the read orientation is incorrect, one read does not map at all, and/or the gap between the reads are of an unexpected size. Although thresholds for gap size and mismatch rates can be changed during read alignment, most studies use inconsistent filters at this stage of the pipeline which can include, but are not limited to, retaining only reads with a certain MAPQ score, retaining proper pairs, or only retaining the primary alignment[31]. Likely, the number of properly paired reads decreases

when the gray fox is mapped to the heterospecific genomes because of structural variations between the genomes, some of which may be smaller insertions and deletions, and some of which may reflect the large chromosomal rearrangements between the species. To exemplify the inconsistency of pipelines and filters applied, we screened five papers which used the domestic dog as a reference genome for analyses of a heterospecific canid taxa and found that four papers did not remove improperly paired reads before variant calling[15,18,20,32], while one paper did remove improperly paired reads[19], and their remaining mapping and filtering choices had almost no overlap. Despite many of these studies filtering on MAPQ, mapping quality is assigned for an individual read by BWA[33], and not as a pair (i.e., the MAPQ is ambiguous to the read being properly paired). GATK automatically filters reads with a MAPQ of 10 or less during variant calling and other filters such as genotype quality (GQ; accuracy of the particular variant call at that site) and QUAL (how confident the variant caller is that there is variation at that site) are also likely to affect the resulting variant calls. Though we do not explore these as part of this work, a systematic investigation of these filters and their downstream impacts on mapping and variant calling, especially in the case of divergent references, is highly needed.

We found that the conspecific gray fox genome provided a more accurate representation of genetic variation, revealing a much larger number of SNPs and a more reliable site frequency spectrum compared to the heterospecific genomes. The increased detection of rare variants indicates that heterospecific references may underestimate the presence of low-frequency alleles, which can have profound implications for downstream analyses, including inference of demographic histories, the estimation of genetic load, and recombination rates. The higher mean SNP depth and lower missing data rates with the gray fox reference further support its superior performance in variant detection, indicating a more accurate and comprehensive alignment that reduces the likelihood of false negatives. Approximately 20% of the variants identified with heterospecific references were incorrectly identified when lifted over to the gray fox genome (Figure S2). Of these, about 5% were misclassified—for example, variants identified as singletons in heterospecific references were non-singleton SNPs in the conspecific reference, and vice versa. The remaining 15% were not recognized as variants at all when mapped to the conspecific genome. The 20% discrepancy likely results from sequence divergence between species, leading to alignment errors or misinterpretation of homologous regions. The higher proportion of matching non-singleton SNPs suggests that common variants are more consistently detected across references, while rare variants are more likely to be missed when using heterospecific genomes. When estimating nucleotide diversity (π), the gray fox reference produced higher values in both populations, indicating that heterospecific genomes can underestimate genetic diversity within populations. The greater difference in π between the western and eastern populations observed with the gray fox genome also suggests that the conspecific reference provides a more sensitive measure of genetic diversity.

Having performed the first demographic inference for the gray fox with a conspecific reference genome, we sought to compare the eastern and western gray fox populations to gain deeper insights into their population histories. These two populations had quite distinct demographic trajectories, with the western population showing overall larger and stable population size with no recent growth. In contrast, the eastern population showed a stable population with some

recent growth. Genetic diversity was significantly higher in the western gray fox population compared to the eastern population, suggesting differing evolutionary pressures and histories in these lineages. Our results differ from previous studies[15] of these populations, which mapped to CanFam3.1 and used PSMC on a single individual to infer the ancient demographic histories of the populations. Specifically, we infer smaller and more stable populations throughout time, with less of a disparity between east and west population sizes. Both our study and the previous study identified the same pattern of the western population being more diverse than the eastern population for most timepoints, though the magnitude of our diversity estimates is much lower than previous estimates[15]. We also did not detect a recent decline in the west, which was found only when mapped to heterospecific references in our analysis as well as in previous work[15].

We utilized both coalescent and SFS-based methods to infer demographic trajectories for both populations, with the hope that one approach may be more robust to the use of a divergent reference. Ultimately, we again observed substantial discrepancies in effective population size estimates when heterospecific reference genomes are used for mapping. When mapped to the heterospecific genomes, the gray fox populations were inferred to have substantially lower effective population sizes, while the gray fox genome produced higher estimates that align more closely with expectations based on species biology and similar studies in other mammals[34]. Importantly, irrespective of population, the inferred demographic trajectory using the divergent dog or Arctic fox genomes were not consistent with the conspecific gray fox throughout most of the inferred history. Further, this pattern was consistent no matter which demographic inference method (SFS based or coalescent) was used (Figure S3). This inconsistency is particularly worrying in the case of species threatened with extinction, because approximately 99% of these species do not have a reference genome[35] and researchers will ultimately be forced to map to a heterospecific reference.

For many species, it is extremely difficult to generate pedigree or gamete-based estimates of the recombination landscape. Thus, researchers will typically use population-based methods to infer recombination maps. One of the commonly used methods is pyrho, which utilizes inferred demographic histories as input to estimate recombination maps for each population. Focusing on the recombination rate estimates inferred with the species-matched reference, we found that landscapes of recombination differed notably between the eastern and western populations, with the west consistently exhibiting higher recombination rates, particularly at chromosome ends. A recombination landscape that is fairly stable with peaks towards the end of the chromosome has been previously observed in dogs[36], and is particularly interesting given that the family *Canidae* has a psuedogenized copy of PRDM9[17,37]. Previous work used this recombination peak to identify the directionality of chromosomes, which has yet to be established for this new reference genome, so future work could take advantage of the result to identify directionality of chromosomes in the gray fox.

Conversely in the case of heterospecific reference genomes, where demographic trajectories were quite discordant from the conspecific reference, the inferred recombination maps and rates differed significantly in both populations. In the eastern population, recombination rates were underestimated with the Arctic fox genome and overestimated with the dog genome compared

to the gray fox genome, while in the west, both heterospecific genomes overestimated recombination rates, with less pronounced differences. The inconsistent direction of bias—where the Arctic fox genome underestimates recombination rates in the eastern population but overestimates them in the western population compared to the gray fox genome—suggests that population-specific patterns may influence how reference bias manifests. This variability can complicate mitigation efforts, as the potential sources of bias differ between populations, making uniform corrections ineffective and highlighting the complexities in addressing reference bias when it varies between populations. In addition, recombination landscapes based on heterospecific genomes showed greater variability, particularly toward chromosome ends, with higher maximum rates than those inferred using the gray fox genome. These patterns highlight how reference genome choice can introduce substantial bias, affecting both the magnitude and distribution of inferred recombination rates.

Potential sources of bias for the inferred recombination rate include chromosomal synteny and spurious singletons. In the context of synteny, previous work has shown that there have been large scale chromosomal rearrangements within the evolutionary history of these species[17]. These karyotype differences have resulted in less than half of the dog and Arctic fox chromosomes being syntenic with gray fox chromosomes (Figure S5; Table S1). Thus, when a recombination map is estimated using a heterospecific genome with karyotypic differences as a reference, we are ultimately disrupting linkage patterns and inferring false recombination events across multiple chromosomes. To investigate this, we examined whether differences in recombination rates between each heterospecific reference and the conspecific reference varied between syntenic and non-syntenic chromosomes but found no difference in the mean recombination rates in either population or genome comparison (Figure S6). The observed inflation in recombination rates could result from disrupted linkage patterns caused by lack of synteny, increased sequence divergence leading to spurious singletons, or an interaction of these factors. Disentangling these contributions is challenging because they are inherently interconnected: synteny disruptions and sequence divergence both affect how variants are identified and mapped, while spurious singletons may arise as an artifact of these processes, obscuring their individual effects. In syntenic regions, pyrho would infer inflated recombination rates when there are spurious singletons and other incorrectly mapped variants being interpreted as evidence of recombination events. The misclassification and misidentification of 20% of the variants as a result of reference bias (Figure S2) is likely what led to skewed estimates of recombination rates, especially in regions of the genome where alignment is poor. The observation of non-syntenic regions and lower reference genome quality introducing bias was also captured in recent work examining an improved olive baboon reference genome[38].

Finally, we examined estimates of F_{ST}, and used the genome-wide distribution of F_{ST} to perform an outlier-scan to detect genes that are potentially under selection. We found that estimates of F_{ST} were higher when using the gray fox genome compared to the values obtained from heterospecific references. We also observed a stronger negative correlation between π and F_{ST} when using the gray fox reference, suggesting that regions of low diversity correspond to areas of high differentiation. The discrepancy among overall π and F_{ST} values across reference genomes is not concerning in terms of magnitude, suggesting that genome-wide averages are

likely sufficient for broadly characterizing diversity within and differentiation between populations. However, studies typically go beyond genome-wide averages, examining variation in F_{ST} across the genome to identify regions of high differentiation, which may indicate the presence of genes under selection or regions contributing to reproductive isolation.

Importantly, we found that the choice of reference genome substantially affected F_{ST} outlier detection. Heterospecific references identified more than twice the number of unique outlier windows compared to the conspecific reference. This discrepancy could be due to alignment issues, where regions not well conserved between species may erroneously appear as regions of high differentiation. The limited overlap of outlier windows between references raises concerns about the reliability of using heterospecific genomes for detecting genomic regions under selection. The distinct GO terms associated with each reference further illustrate how reference genome choice shapes the inferred biological processes underlying F_{ST} outliers. For instance, unique GO terms in the Arctic fox reference emphasized cellular signaling and developmental regulation, while Canfam4 highlighted growth and endocrine response, and the gray fox reference captured nervous system function and immune modulation. These differences underscore the biological biases introduced by reference genome choice. In sum, using a heterospecific reference genome would inflate false positives for both the F_{ST} outliers, genes, and biological pathways that may be under selection. Thus, using a conspecific reference would certainly provide more accurate insights into the biological processes influencing genetic differentiation within and between the species.

Our work provides important insights into the effects of reference bias, however, there still exists some limitations. One specific limitation is that the gray fox reference genome used here comes from an individual from the east coast. We see that even human populations with smaller divergence estimates are affected by reference bias[39]. Future work could either focus on correcting for this bias computationally or generating a west coast gray fox reference genome. The second caveat is that we are quantifying reference bias in *Canidae*, which is notable for being a mammalian clade with karyotype shifts. Despite these shifts, *Canidae* species are relatively young, which likely moderates the effects of reference bias compared to more divergent species. Non-mammalian taxa exhibiting greater evolutionary divergence are likely to experience even greater impacts of reference bias. Since non-mammalian species make up the bulk of the 99% of species without reference genomes, this highlights the broader importance of addressing reference bias. At the same time, karyotype differences are common even between closely related species, making *Canidae* a useful system for studying the interplay between karyotype shifts and reference bias.

This study highlights the critical impact of reference genome choice on population genomic inferences and emphasizes the value of conspecific genomes in uncovering accurate evolutionary histories. Quantifying reference bias using *Canidae* has implications for some of the world's most endangered species, such as the Ethiopian wolf (*Canis simensis*) and the African wild dog (*Lycaon pictus*), and more common species, like the gray fox and coyote (*Canis latrans*). Reference bias may also affect our understanding of historic hybridization and introgression, a phenomenon common in the canid clade[32,40]. Furthermore, genomic data are

increasingly being used to inform conservation management plans. Measures such as adaptive capacity[41] and differential adaptations between populations[42] are being used to make recommendations regarding translocations and rewilding. Our selection scan results suggest that these inferences can be skewed by reference bias when divergent reference genomes are used.

Our results underscore the necessity of using conspecific reference genomes in conservation genetics and evolutionary studies, particularly for accurately understanding divergence and diversity in non-model organisms. However, we acknowledge that developing high-quality reference genomes for all species is not always feasible due to resource constraints. In such cases, careful consideration should be given to the phylogenetic proximity and degree of synteny of the available reference genomes. Our work demonstrates that even closely related heterospecific references may not adequately capture the genetic landscape of the target species. Future studies should focus on improving reference genome assemblies for non-model organisms, leveraging advancements in sequencing technologies and assembly algorithms. Additionally, methodologies that are less reference-dependent, such as reference-free variant calling or pangenomic approaches, may help mitigate some of the biases introduced by heterospecific references.

We demonstrate that using a conspecific reference genome enhances the inference of population size histories and recombination rates, improves the detection of genetic variation, provides more accurate estimates of nucleotide diversity and genetic differentiation, and influences the biological interpretation of genomic data. These findings highlight the broader implications for genomic research in other non-model organisms and stress the need for continued efforts in generating high-quality, species-specific genomic resources and suggest that results using highly divergent reference genomes should be interpreted with caution.

Methods

Whole genome sequence data and reference genomes

To investigate the impact of reference bias on population genomic inference of diversity, demography, selection, and recombination in gray foxes, we obtained whole-genome resequencing data for 41 gray foxes sampled across North America (Figure S1) from Preckler-Quisquater et al.[15], in which the data were mapped to the domestic dog reference genome CanFam3.1[21]. We re-mapped the data to three reference genomes: (1) Arctic fox[23], (2) a more contiguous updated version of the dog genome CanFam4[22], and (3) our recently published gray fox reference genome[17], based on an individual sampled in Vermont (Figure 1A). We identified syntenic and non-syntenic chromosomes between species using information from Graphodatsky et al.[43] and through alignments generated in Armstrong et al.[17]. All putatively syntenic chromosomes from Graphodatsky et al.[43] were then manually interrogated using the genomic alignments from Armstrong et al.[17]. Any putatively syntenic chromosome that mapped to more than one chromosome in either species was removed and considered non-syntenic to be as conservative as possible.

Mapping and variant calling

We mapped whole-genome data from gray fox individuals to each of three reference genomes (see above) using identical pipelines. Sequence data was mapped using NVIDIA Parabricks v4.1.1.1[44] *fq2bam* with the provided reference and default parameters. We estimated depth and other mapping statistics using Mosdepth v0.3.3[45] and SAMtools v1.16.1 *flagstat*, respectively. We tested for differences in these measures using pairwise.wilcox.test with p.adjust.method = "none" in R v4.2.1[46]. Resulting BAM files were individually run through Parabricks *HaplotypeCaller* with the '--gvcf' flag in order to emit both variant and invariant site calls. Subsequently, GATK v4.1.4.1[47] *GenomicsDBImport* was used to import the single sample VCFs prior to joint genotyping with the '-all-sites' flag to retain both variant and invariant sites. Autosomal chromosomes were provided as intervals, excluding sex chromosomes and unlinked scaffolds for downstream analyses. Finally, we ran Parabricks *GenotypeGVCFs* with the '--all-sites' flag to produce final gVCF files. Files were combined using BCFtools v1.16 *concat* to create genome-wide variant calls.

Genome annotations

We produced mappability scores for each genome using GenMap v1.3.0[48]. Mappability scores are used to assess the uniqueness of kmers in the genome and identify regions which are repetitive and cause errors during mapping and variant calling. We used the filterGM.rb script from Armstrong & Campana[49] to generate a file of sites with a mappability score < 1.

We also generated repetitive element annotations for each genome. Briefly, we ran TETools v1.7[50] on each genome assembly as in Armstrong et al.[17]. Final output files were converted to bed files using the RM2bed.rb script and subsequently combined with the mappability scores file using the simplify_bed.rb script[49].

Genome annotations were previously generated for both versions of the dog genome, however, annotations have not been previously generated for the gray fox. Ideally, a combination of evidence from RNA and closely related species annotations would be used to annotate the gray fox genome, but RNA data has not been generated for this species. As such, we used the GALBA v1.0.9[51] program to generate draft annotations for the gray fox. Though these annotations are likely imperfect, since we are only using annotation information to remove regions which may be evolving non-neutrally, these are sufficient for our purposes of exclusion. We downloaded protein files from the arctic fox (GCF_018345385.1[23]), domestic dog (GCF_011100685.1[22]), brown bear (GCF_023065955.2[52]), and the most recent (at the time) human genome assembly (GCF_009914755.1[53]). These files were provided as protein evidence to GALBA using the –prot_seq flag and run using AUGUSTUS v3.3.2[54] which was run with default parameters.

Variant filtering

Both repetitive and low-mappability regions were filtered using the BCFtools *view* command and providing the described bed file after the '-T ^' flags. We then assessed quality statistics using BCFtools *query* with the flags '-f' and pulled the statistics for allelic number (AN) and depth (DP). We calculated the mean and interquartile range for allelic number and depth for these statistics for each genome. Subsequently, we used the BCFtools *filter* command with the '-i' flag to include variants which had more than 90% of the maximum AN value (2* number of

samples), a quality score greater than 30 ('QUAL >= 30'), and a depth that was greater than the 25% percentile and below 1.5 times the mean depth, approximately. Finally, only biallelic sites were retained using BCFtools view with the '-M 2' flags. For specific filters and values see Github https://github.com/ellieearmstrong/Gray_Fox_2023/tree/main/filtering.

Sample selection

We performed a principal component analysis (PCA) on SNP genotype data including all 41 individuals using the SNPRelate v1.36.1 package in R v4.3.2[46]. We first converted the VCF file to a genomic data structure file using the snpgdsVCF2GDS function. PCA was conducted with the snpgdsPCA function on genotype data with no missing values, generating eigenvectors for each sample and eigenvalues representing the variance explained by each principal component. Based on the PCA results (Figure S1), we subsampled the dataset to exclude hybrids, which are known to occur in the southwest sampling sites. To achieve this, we selected the westernmost samples in PC space, prioritizing the inclusion of the high-coverage sample, and matched this selection by choosing the same number of easternmost samples.

Computing genetic diversity (π) and differentiation (F_{ST})

Pairwise genetic diversity (π) within each population and genetic differentiation (F_{ST}) between populations were computed using VCFtools[55] v1.14, including six individuals from the east and six individuals from the west. We calculated π using a gVCF file containing all genome sites—both variant and invariant—and kept only sites without missing data for each population. F_{ST} was calculated using a VCF file with only variant sites, retaining sites with no missing data across populations. We computed both summary statistics in 50 kb windows. We examined the relationship between genetic diversity and differentiation for each population and reference genome using a Kendall's rank correlation test, given the exponential-like distribution of nucleotide diversity.

Quantifying genetic variation

We used VCFtools[55] to obtain raw allele counts at all biallelic variant sites, which we summarized to get the total number of variants as well as the number of singletons, i.e., rare variants where an allele is only found once. To calculate an SFS for each population based on each reference genome, we used VCFtools to obtain allele counts at all fully covered (i.e., no missing data) biallelic sites, which we summarized to get a folded SFS. Data summaries were performed using the tidyverse package in R v. 4.3.2[46]. Site frequency spectra were generated using 6 individuals from the east and 6 individuals from the west for the conspecific and heterospecific reference genomes.

SNPs with missing data were removed from these analyses. There were 3,496,241 SNPs in the East and 2,669,112 SNPs in the West out of the 26,180,875 total SNPs that were removed due to missing data with Gray fox reference genome; 11,031,507 SNPs in the East and 6,846,230 SNPs in the West out of the 43,583,377 total SNPs that were removed due to missing data with CanFam4 reference genome; and 11,526,220 SNPs in the East and 7,176,780 SNPs in the West out of the 45,042,040 total SNPs that were removed due to missing data with Arctic fox reference genome.

Demographic inference

Demography was inferred using 6 individuals from the east and 6 individuals from the west with smc++[56] v1.15.4, MSMC2[57] v2.1.4, and Stairway Plot 2 v2.1.2[58]. Smc++ combines the SFS with linkage disequilibrium (LD) information in coalescent hidden Markov models (HMM), MSMC2 uses a coalescent HMM without the SFS, and Stairway Plot 2 relies solely on the SFS.

For smc++, we parsed VCFs to generate input files for each autosomal chromosome using the vcf2smc command in smc++. Because smc++ is not able to distinguish regions of missing data from very long runs of homozygosity, positions of regions with highly repetitive sequence and/or low mappability that were masked prior to variant calling were provided in a bed file and marked as missing data using the -m option in the outputted smc++ data sets. Positions of genic regions were obtained from annotations and similarly masked such that demographies were inferred using only putatively neutral regions of the genome. Further, smc++ calculates the SFS by using a single “distinguished individual” selected from the pool of samples against which all other samples are compared, which may impact demographic inference depending on the choice of distinguished individual. We therefore created multiple input files by varying the identity of the distinguished individual and treating the remaining samples from each population as undistinguished. Input files were combined to generate a composite likelihood estimate by running the estimate command for fitting a population size history to the data six separate times, once for each combination of population and reference genome. Runs were conducted assuming a per-site per-generation mutation rate of $4.5e{-}9$[59] and 25 EM iterations, with estimates restricted between 2,500 and 500,000 generations since present. A thinning parameter of 1,792, calculated as $1{,}000 \times \log(6)$ for the six individuals in our analysis, was used as recommended in the smc++ user guide to control the frequency of conditional SFS emission and incorporate information from the undistinguished portion of the sample. A spline representation of population size history was fit using 12 knots to allow for sufficient flexibility while avoiding over-smoothing. We used a generation time of two years[14] to convert the output from coalescent units to units of time.

For Stairway Plot 2, we used the folded site frequency spectra that were generated for each population and reference genome above as input. We used all bins of the SFS, total callable sites incorporated the monomorphic sites, a per-site per-generation mutation rate of $4.5e{-}9$[59], a generation time of two years[14], and 67% of the data was used for training. Stairway Plot 2 also implements a check for over or underfitting using various breakpoints. The breakpoints tested were (2, 5, 7, 10) and the best breakpoint value was selected as one that minimized the log-likelihood.

For MSMC2, we first used BCFtools to remove sites with missing data from the gVCF file for each population and reference genome, generated a BED file of sufficiently covered sites for masking, then split files to generate a mask for each chromosome and a gVCF for each sample and chromosome. We then generated input files using generate_multihetsep.py in the MSMC-tools repository (https://github.com/stschiff/msmc-tools), including a negative mask to remove regions with low mappability scores. We estimated population size histories by specifying pair indices (-I) for each individual, to avoid pairs of haplotypes from different individuals, as recommended for unphased genomes.

Recombination rate estimation

Recombination rates were estimated using pyrho v0.1.0[3], which takes in the demography from smc++ and linkage information from unphased data to estimate population-specific fine-scale recombination rates per generation[3]. Using pyrho, we first precomputed likelihood tables under the demographic models we inferred from smc++ to account for population size fluctuations across time and the mutation rate (4.5e−9[59]). We then tested multiple sets of hyperparameters to produce reasonable data from the optimization function. Parameters for block penalty, which controls the smoothness of the resulting map, and window size, the amount of bp before a SNP is ignored, were selected based on their minimization of log(l_2). Block penalty parameters tested were (25, 50, 100, 150, 200, 250) and window sizes tested were (25, 50, 100, 150). After hyperparameter selection (Table 1), we averaged the per-base recombination rate inferred by pyrho into 50 kb windows and converted the values to units of centi-Morgan per megabase by multiplying by 1e8 (100 cM per expected crossover multiplied by 1e6 bases per mb).

Table 1. Pyrho hyperparameter settings for each reference genome and both populations.

	Block Penalty		Window Size (bp)	
	East	West	East	West
Gray fox	25	25	50	25
Canfam4	25	50	50	50
Arctic fox	50	25	150	50

Identifying F_{ST} outliers and conduction functional enrichment analyses

Outliers were defined as windows with F_{ST} values exceeding three standard deviations above the mean for each reference genome. To compare outlier regions across references, we converted window positions between each heterospecific reference and the gray fox reference (described below). Outliers were categorized based on whether they were shared across references or unique to a specific reference. To identify protein-coding genes within the outlier windows, we used Arctic fox and Canfam4 gene annotations, and the draft annotations we generated for the gray fox (described above). We used the GenomicRanges[60] v1.54.1 package in R v. 4.3.2[46] to determine which genes fell within the 50 kb windows of each reference genome. Foreground genes were defined as those located within the outlier windows, while background genes included those located outside the outlier windows. We then conducted a functional enrichment analysis for gene ontology (GO) biological processes using Enrichr[61] to identify processes significantly overrepresented in the foreground gene set compared to the background set. We used the ComplexUpset[62] v1.3.3 R package to create upset plots showing the number of identified genes from each reference and their intersection.

Converting coordinates between reference genomes

We performed whole genome alignments between the two heterospecific genomes and the gray fox genome using minimap2[63] v2.14 with parameters --cs -x asm20. We converted the pairwise alignment format computed by minimap2 to a chain file using paf2chain[64], then used the Crossmap[65] v0.7.0 *bed* command to convert SNP positions and the *region* command, with the default (-r) value of 0.85 for the minimum ratio of bases that must remap, to convert window positions in the heterospecific genomes to their corresponding positions in the gray fox genome.

Acknowledgements

We want to thank Joshua Schraiber and Jeffrey Spence for insightful and helpful conversations during this work. J.A.M was supported by the startup funds from Dornsife College of Letters, Arts and Sciences through the Department of Quantitative and Computational Biology and the USC WiSE Gabilan Assistant Professorship. M.G. was supported by USC Provost's Fellowship.

Data accessibility and benefit-sharing

GitHub code for filtering can be found here https://github.com/ellieearmstrong/Gray_Fox_2023/tree/main/filtering and analyses can be found at https://github.com/makopyan/fox/.

Author contributions

M.A. and J.A.M. conceived of the project. J.A.M., M.A., and M.G. performed data processing and analyses. E.A. conducted mapping and generated variant call files used for analyses. J.A.M., M.A., E.A., and M.G. wrote the manuscript. All authors approved of the final manuscript.

Supplementary Figures

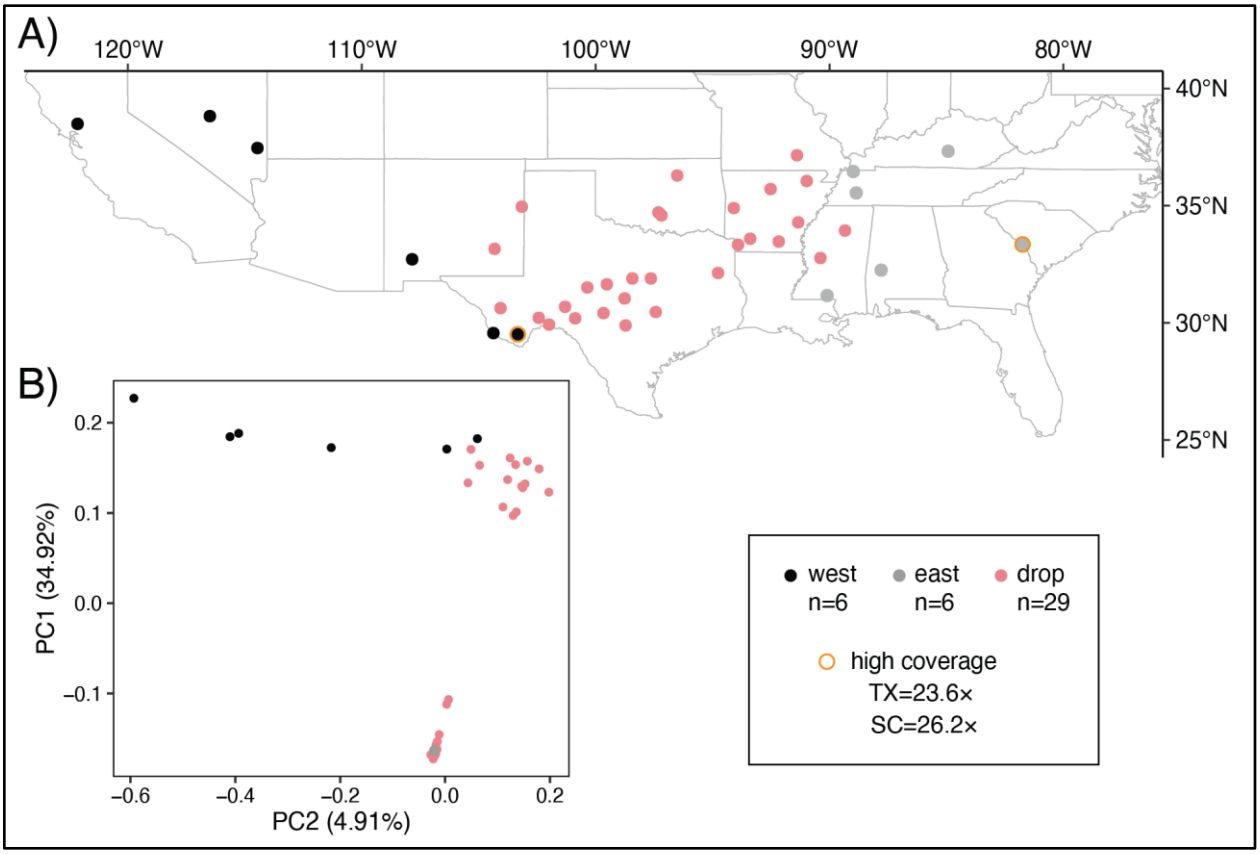

Figure S1. A) *Sampling localities of WGS data for n=41 gray foxes obtained from Preckler-Quisquater et al.*[15] *with gold borders highlighting high-coverage samples.* ***B)*** *Principal components 1 and 2 of genome-wide variation among individuals. In both panels, black and gray points represent the 12 samples included in our analysis, and the pink points represent the excluded samples, which include admixed individuals.*

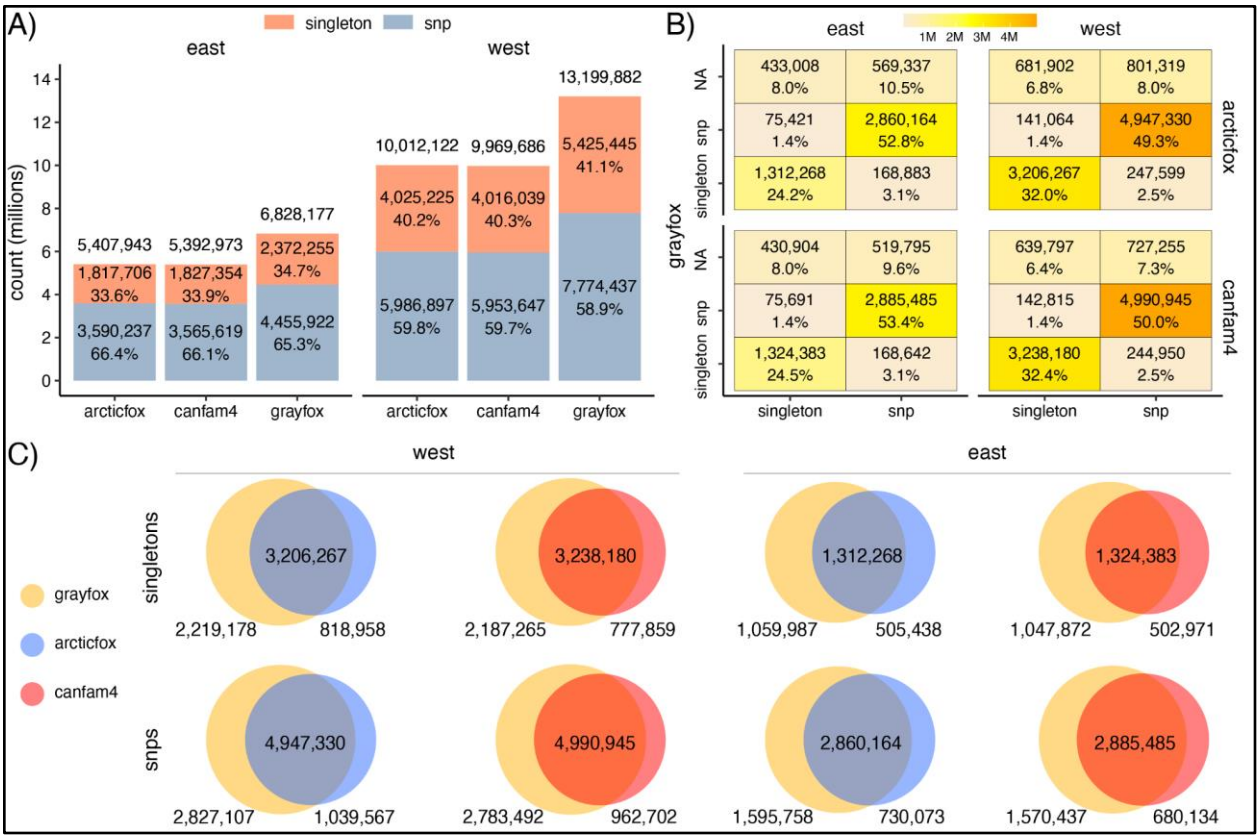

Figure S2. Number and classification of variants by reference genome and population. A) *Bar chart of variants identified in each population with each reference genome, indicating counts and percentages of singletons (orange) and non-singleton SNPs (blue). Values above each bar indicate the total number of SNPs identified in each case. We identified nearly twice the number of total SNPs in the west compared to the east population with all reference genomes. The conspecific gray fox reference yielded more SNPs in both populations compared to the heterospecific references, especially in the west.* ***B)*** *Heatmap showing the distribution of variants identified in east and west populations (top labels) using heterospecific reference genomes (right labels: arctic fox and canfam4). The x-axis distinguishes between singleton SNPs and non-singleton SNPs in the heterospecific reference, while the y-axis categorizes variants based on whether they match the same type of variant or are absent in the gray fox (conspecific) reference genome. The numbers in each cell represent the total number of variants, followed by the percentage of the total variants in the respective category. The color gradient reflects the variant counts in millions, ranging from light beige (lower counts) to dark orange (higher counts).* ***C)*** *Venn diagram indicating the overlap in singletons (top row) and non-singleton SNPs (bottom row) identified in east and west populations across three reference genomes. Overlapping regions indicate variants identified in both reference genomes, with the total matching variant count shown in the center of each diagram. Non-overlapping areas represent variants unique to each reference genome. The numbers adjacent to each diagram denote the count of variants in each set, emphasizing shared and unique SNPs or singletons among the three reference genomes for each population.*

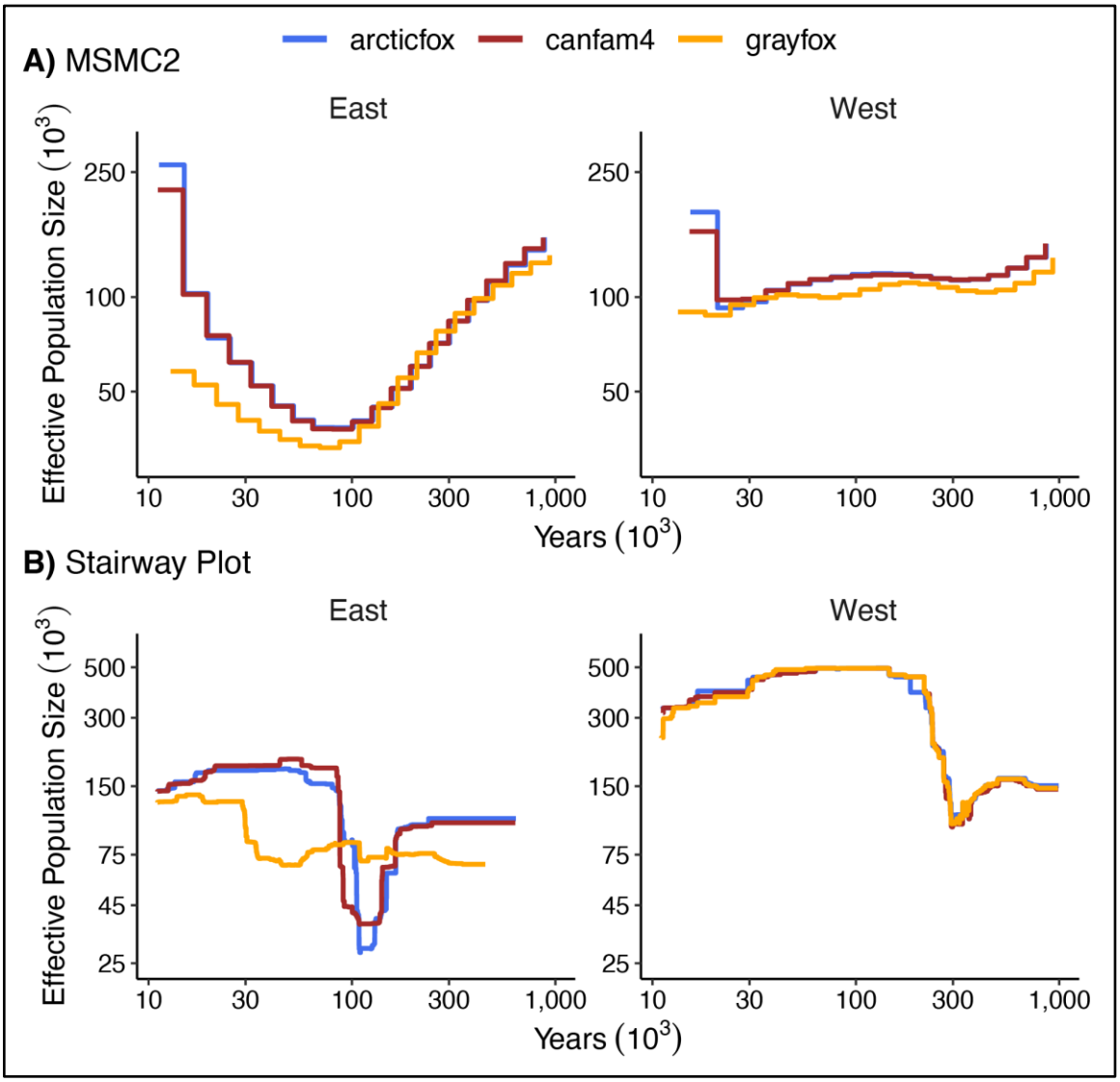

Figure S3. *Inferred effective population sizes (y-axis) and years from present (x-axis) reveal discordant demographic histories of foxes resolved using the species-matched (gold) and heterospecific genomes (red and blue) in the east (left) and the west (right) using* ***A) MSMC2*** *and* ***B) Stairway Plot****.*

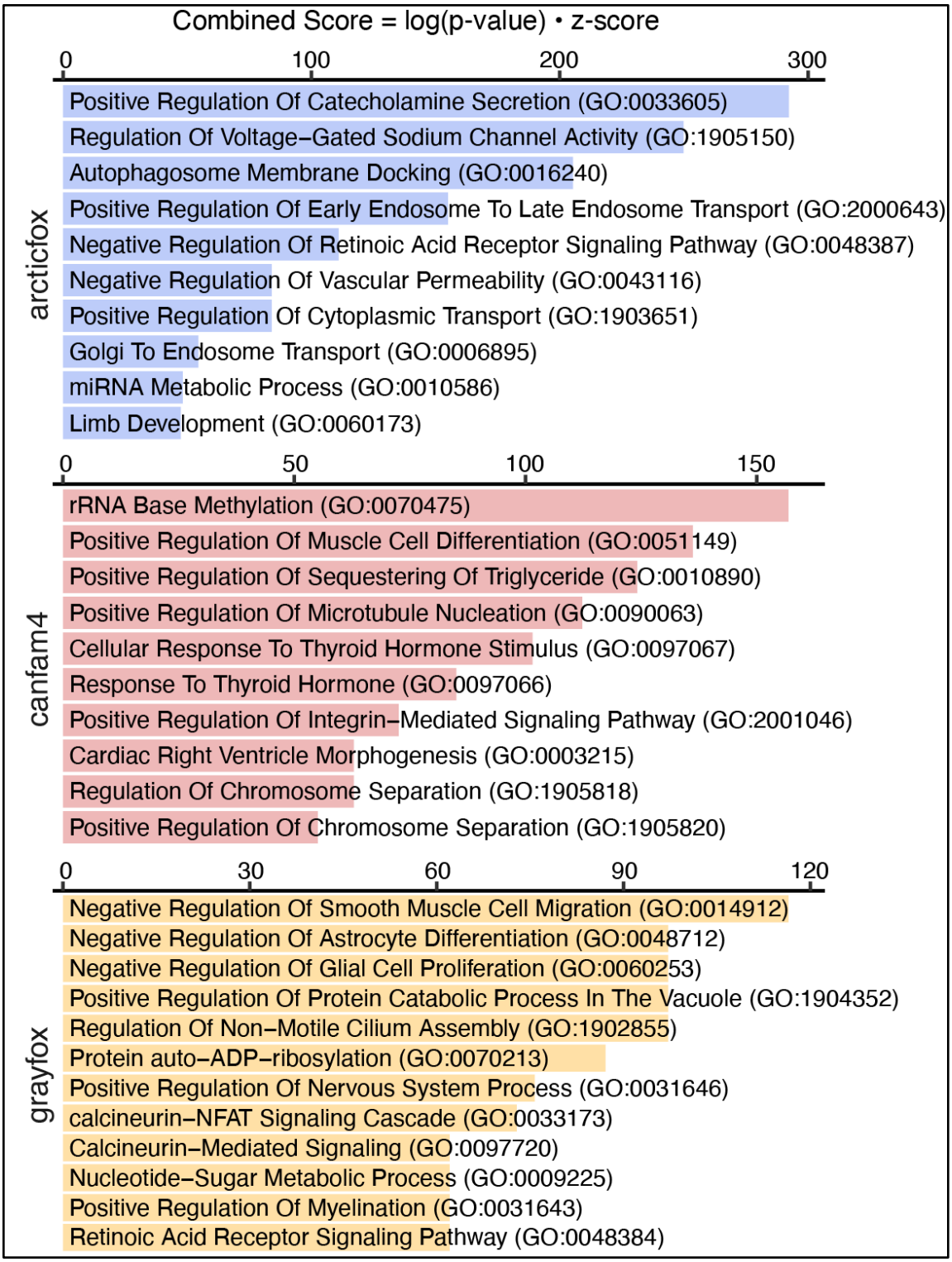

Figure S4. Top 10 unique GO biological process terms identified for each reference genome, highlighting distinct thematic focuses. *The unique terms with the highest combined scores are shown for arctic fox (blue), Canfam4 (red), and gray fox (yellow). Each term is listed alongside its GO identifier, with the combined score on the x-axis calculated as the log(p-value) multiplied by the z-score.*

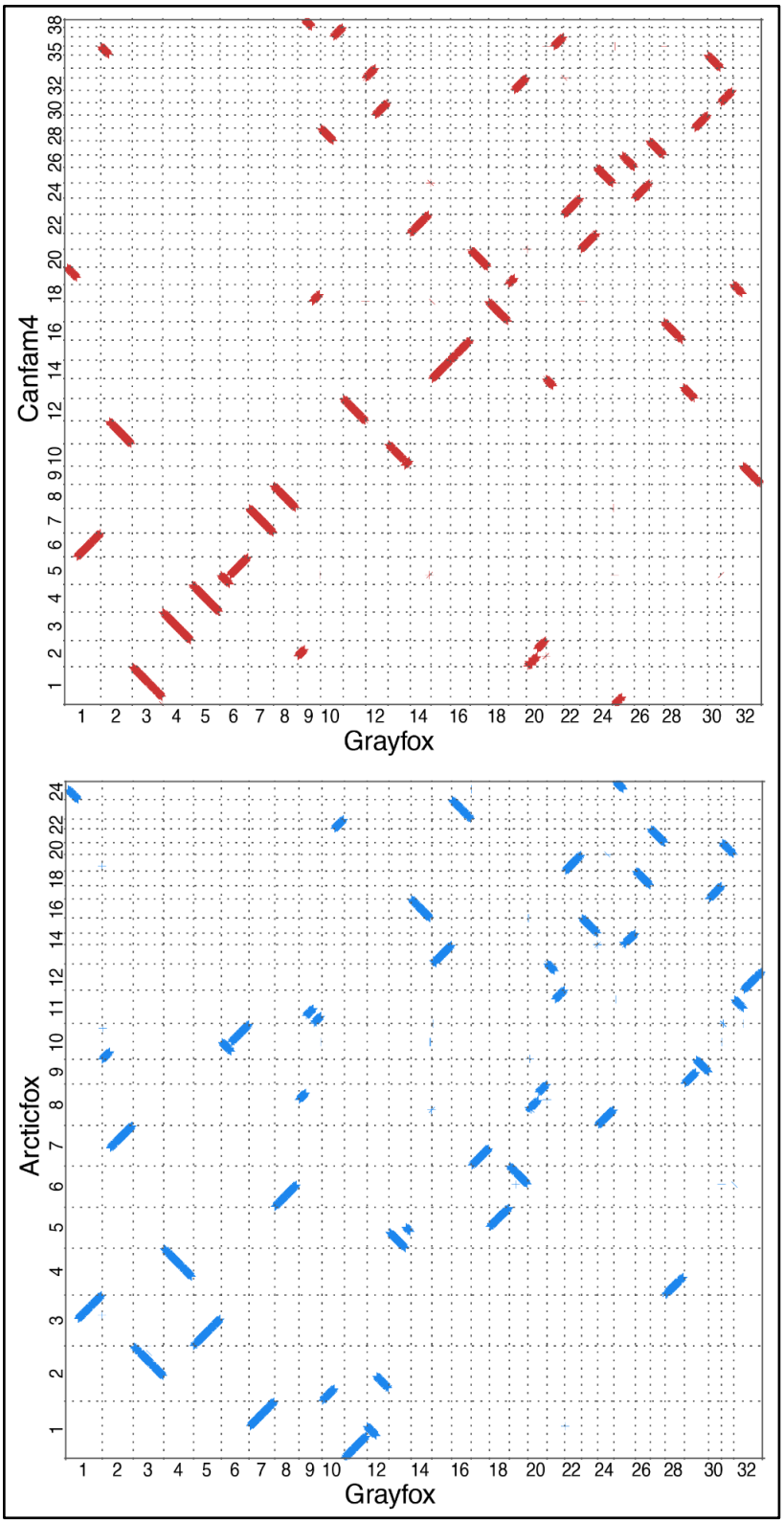

Figure S5. *Dot plot of pairwise whole-genome alignment between the gray fox genome (x-axis) and the heterospecific reference genomes (y-axis) Canfam4 (A) and Arctic fox (B) showing synteny is not conserved, with evidence of chromosomal fusions and fissions.*

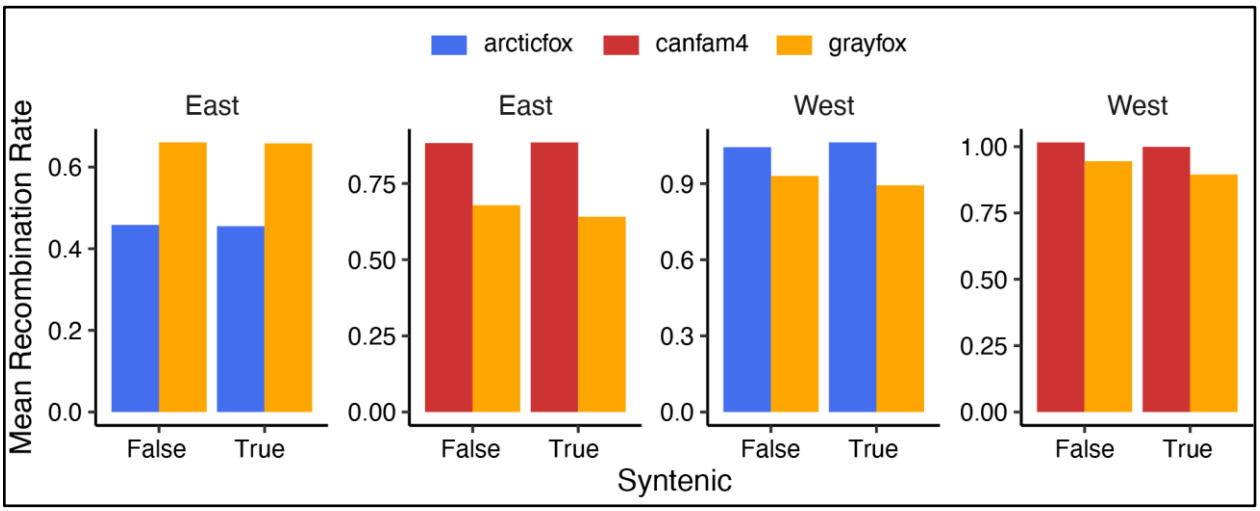

Figure S6. *Mean recombination rates for syntenic vs non-syntenic chromosomes between reference genomes.*

Table S1. *Syntenic chromosomes between gray fox and* ***A)*** *dog and* ***B)*** *Arctic fox genomes.*

A)

Canfam4	Gray fox
3	4
4	5
7	7
8	8
10	13
12	11
14	15
15	16
16	28
17	18
20	17
21	23
24	26
27	27
31	31
34	30

B)

Arctic fox	Gray fox
13	15
15	23
16	14
17	30
18	26
19	22
20	31
21	27
23	16

References

1. Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019). Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 20, 1*17*. https://doi.org/*10*.1*186*/s1*305*9-019-1720-5.
2. Salmona, J., Heller, R., Lascoux, M., and Shafer, A. (2017). Population genomics, concepts, approaches and applications. Popul. Genom., 511–537. https://doi.org/10.1007/138*36*_2017_1.
3. Spence, J.P., and Song, Y.S. (2019). Inference and analysis of population-specific fine-scale recombination maps across *26* diverse human populations. Sci. Adv. 5, eaaw9206. https://doi.org/10.1126/sciadv.aaw9206.
4. Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. *15*, e1008302. https://doi.org/10.1371/journal.pgen.1008302.
5. Gopalakrishnan, S., Ebenesersdóttir, S.S., Lundstrøm, I.K.C., Turner-Walker, G., Moore, K.H.S., Luisi, P., Margaryan, A., Martin, M.D., Ellegaard, M.R., Magnússon, Ó. þ., et al. (20*22*). The population genomic legacy of the second plague pandemic. Curr. Biol. *32*, 4743-4751.e6. https://doi.org/10.10*16*/j.cub.2022.09.0*23*.
6. Günther, T., and Schraiber, J.G. (20*24*). Estimating allele frequencies, ancestry proportions and genotype likelihoods in the presence of mapping bias. bioRxiv, 2024.07.01.601500. https://doi.org/10.1101/2024.07.01.601500.
7. Nevado, B., Ramos-Onsins, S.E., and Perez-Enciso, M. (20*14*). Resequencing studies of nonmodel organisms using closely related reference genomes: optimal experimental designs and bioinformatics approaches for population genomics. Mol Ecol 23, 1764–1779. https://doi.org/10.1111/mec.12693.
8. Gopalakrishnan, S., Castruita, J.A.S., Sinding, M.-H.S., Kuderna, L.F.K., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., Marques-Bonet, T., et al. (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genom. 18, *49*5. https://doi.org/10.1186/s1*28*64-017-3883-3.
9. Armstrong, E.E., Taylor, R.W., Miller, D.E., Kaelin, C.B., Barsh, G.S., Hadly, E.A., and Petrov, D. (2020). Long live the king: chromosome-level assembly of the lion (Panthera leo) using linked-read, Hi-C, and long-read data. BMC Biol. 18, 3. https://doi.org/10.1186/s12915-019-0*734*-5.
10. Thorburn, D.J., Sagonas, K., Binzer-Panchal, M., Chain, F.J.J., Feulner, P.G.D., Bornberg-Bauer, E., Reusch, T.B.H., Samonte-Padilla, I.E., Milinski, M., Lenz, T.L., et al. (2023). Origin matters: Using a local reference genome improves measures in population genomics. Mol. Ecol. Resour. 23, 1706–1723. https://doi.org/10.1111/1755-0998.13838.
11. Maurstad, M.F., Hoff, S.N.K., Cerca, J., Ravinet, M., Bradbury, I., Jakobsen, K.S., Præbel, K., and Jentoft, S. (2024). Reference genome bias in light of species-specific chromosomal reorganization and translocations. bioRxiv, 2024.06.28.599671. https://doi.org/10.1101/2024.06.28.599671.
12. Rick, J.A., Brock, C.D., Lewanski, A.L., Golcher-Benavides, J., and Wagner, C.E. (2023). Reference genome choice and filtering thresholds jointly influence phylogenomic analyses. Syst. Biol. 73, 76–101. https://doi.org/10.1093/sysbio/syad065.
13. Prasad, A., Lorenzen, E.D., and Westbury, M.V. (2022). Evaluating the role of reference-

genome phylogenetic distance on evolutionary inference. Mol. Ecol. Resour. 22, *45*–55. https://doi.org/10.1111/1755-0998.13457.
14. Goddard, N.S., Statham, M.J., and Sacks, B.N. (2015). Mitochondrial analysis of the most basal canid reveals deep divergence between eastern and western North American gray foxes (Urocyon spp.) and ancient roots in Pleistocene California. PLoS ONE 10, e0136329. https://doi.org/10.1371/journal.pone.0136329.
15. Preckler-Quisquater, S., Kierepka, E.M., Reding, D.M., Piaggio, A.J., and Sacks, B.N. (2023). Can demographic histories explain long-term isolation and recent pulses of asymmetric gene flow between highly divergent grey fox lineages? Mol. Ecol. 32, 5323–5337. https://doi.org/10.1111/mec.17105.
16. Wayne, R.K., Nash, W.G., and O'Brien, S.J. (1987). Chromosomal evolution of the Canidae. Cytogenet. Genome Res. *44*, 123–133. https://doi.org/10.1159/000132356.
17. Armstrong, E.E., Bissell, K.L., Fatima, H.S., Heikkinen, M.A., Jessup, A., Junaid, M.O., Lee, D.H., Lieb, E.C., Liem, J.T., Martin, E.M., et al. (2024). Chromosome-level assembly of the gray fox (Urocyon cinereoargenteus) confirms the basal loss of PRDM9 in Canidae. G3: Genes, Genomes, Genet. 14, jkae034. https://doi.org/10.1093/g3journal/jkae034.
18. Robinson, J.A., Vecchyo, D.O.-D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D., Lohmueller, K.E., and Wayne, R.K. (2016). Genomic flatlining in the endangered island fox. Curr Biol 26, 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062.
19. Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E., and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Curr. Biol. 28, 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066.
20. Mooney, J.A., Marsden, C.D., Yohannes, A., Wayne, R.K., and Lohmueller, K.E. (2022). Long-term small population size, deleterious variation, and altitude adaptation in the Ethiopian wolf, a severely endangered canid. Mol. Biol. Evol. *40*, msac*277*. https://doi.org/10.1093/molbev/msac277.
21. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819. https://doi.org/10.1038/nature04338.
22. Wang, C., Wallerman, O., Arendt, M.-L., Sundström, E., Karlsson, Å., Nordin, J., Mäkeläinen, S., Pielberg, G.R., Hanson, J., Ohlsson, Å., et al. (2021). A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. Commun. Biol. 4, 185. https://doi.org/10.1038/s42003-021-01698-x.
23. Peng, Y., Li, H., Liu, Z., Zhang, C., Li, K., Gong, Y., Geng, L., Su, J., Guan, X., Liu, L., et al. (2021). Chromosome-level genome assembly of the Arctic fox (Vulpes lagopus) using PacBio sequencing and Hi-C technology. Mol. Ecol. Resour. 21, 2093–2108. https://doi.org/10.1111/1755-0998.13*397*.
24. Lewin, H.A., Richards, S., Aiden, E.L., Allende, M.L., Archibald, J.M., Bálint, M., Barker, K.B., Baumgartner, B., Belov, K., Bertorelle, G., et al. (2022). The Earth BioGenome Project 2020: Starting the clock. Proc. Natl. Acad. Sci. *119*, e2115635118. https://doi.org/10.1073/pnas.2115635118.
25. Galla, S.J., Forsdick, N.J., Brown, L., Hoeppner, M., Knapp, M., Maloney, R.F., Moraga, R.,

Santure, A.W., and Steeves, T.E. (2018). Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. Genes 10, 9. https://doi.org/10.3390/genes10010009.
26. Minias, P., Dunn, P.O., Whittingham, L.A., Johnson, J.A., and Oyler-McCance, S.J. (2019). Evaluation of a chicken 600K SNP genotyping array in non-model species of grouse. Sci. Rep. 9, 6407. https://doi.org/10.1038/s41598-019-42885-5.
27. Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., Ruiz-Hueso, P., and González-Candelas, F. (2021). One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. PLoS Comput. Biol. 17, e1008678. https://doi.org/10.1371/journal.pcbi.1008678.
28. Yang, X., Lee, W.-P., Ye, K., and Lee, C. (2019). One reference genome is not enough. Genome Biol. 20, 104. https://doi.org/10.1186/s13059-019-1717-0.
29. Miga, K.H., and Wang, T. (2021). The Need for a Human Pangenome Reference Sequence. Annu. Rev. Genom. Hum. Genet. 22, 1–22. https://doi.org/10.1146/annurev-genom-120120-081921.
30. Wu, X., Heffelfinger, C., Zhao, H., and Dellaporta, S.L. (2019). Benchmarking variant identification tools for plant diversity discovery. BMC Genom. 20, 701. https://doi.org/10.1186/s12864-019-6057-7.
31. Hemstrom, W., Grummer, J.A., Luikart, G., and Christie, M.R. (2024). Next-generation data filtering in the genomics era. Nat. Rev. Genet. 25, 750–767. https://doi.org/10.1038/s41576-024-00738-6.
32. Chavez, D.E., Gronau, I., Hains, T., Dikow, R.B., Frandsen, P.B., Figueiró, H.V., Garcez, F.S., Tchaicka, L., Paula, R.C. de, Rodrigues, F.H.G., et al. (2022). Comparative genomics uncovers the evolutionary history, demography, and molecular adaptations of South American canids. Proc. Natl. Acad. Sci. 119, e2205986119. https://doi.org/10.1073/pnas.2205986119.
33. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
34. Wilder, A.P., Supple, M.A., Subramanian, A., Mudide, A., Swofford, R., Serres-Armero, A., Steiner, C., Koepfli, K.-P., Genereux, D.P., Karlsson, E.K., et al. (2023). The contribution of historical processes to contemporary extinction risk in placental mammals. Science *380*, eabn5856–eabn5856. https://doi.org/10.1126/science.abn5856.
35. Paez, S., Kraus, R.H.S., Shapiro, B., Gilbert, M.T.P., Jarvis, E.D., Group, V.G.P.C., Al-Ajli, F.O., Ceballos, G., Crawford, A.J., Fedrigo, O., et al. (2022). Reference genomes for conservation. Science *377*, 364–366. https://doi.org/10.1126/science.abm8127.
36. Campbell, C.L., Bhérer, C., Morrow, B.E., Boyko, A.R., and Auton, A. (2016). A pedigree-based map of recombination in the domestic dog genome. G3: Genes, Genomes, Genet. 6, 3517–3524. https://doi.org/10.1534/g3.116.034678.
37. Axelsson, E., Webster, M.T., Ratnakumar, A., Consortium, T.L., Ponting, C.P., and Lindblad-Toh, K. (2012). Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. 22, 51–63. https://doi.org/10.1101/gr.124123.111.
38. Batra, S.S., Levy-Sakin, M., Robinson, J., Guillory, J., Durinck, S., Vilgalys, T.P., Kwok, P.-Y., Cox, L.A., Seshagiri, S., Song, Y.S., et al. (2020). Accurate assembly of the olive baboon (Papio anubis) genome using long-read and Hi-C data. GigaScience 9, giaa134. https://doi.org/10.1093/gigascience/giaa134.

39. Brandt, D.Y.C., Aguiar, V.R.C., Bitarello, B.D., Nunes, K., Goudet, J., and Meyer, D. (2015). Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 Genomes Project phase I data. G3: Genes, Genomes, Genet. 5, 931–941. https://doi.org/10.1534/g3.114.015784.
40. Gopalakrishnan, S., Sinding, M.-H.S., Ramos-Madrigal, J., Niemann, J., Castruita, J.A.S., Vieira, F.G., Carøe, C., Montero, M. de M., Kuderna, L., Serres, A., et al. (2018). Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr. Biol. 28, 3441-3449.e5. https://doi.org/10.1016/j.cub.2018.08.041.
41. Funk, W.C., Forester, B.R., Converse, S.J., Darst, C., and Morey, S. (2019). Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. Conserv. Genet. 20, 115–134. https://doi.org/10.1007/s10592-018-1096-1.
42. Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R., Oakeshott, J., Weeks, A., Joseph, L., et al. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. Clim. Chang. Responses 2, 1. https://doi.org/10.1186/s40665-014-0009-x.
43. Graphodatsky, A.S., Perelman, P.L., Sokolovskaya, N.V., Beklemisheva, V.R., Serdukova, N.A., Dobigny, G., O'Brien, S.J., Ferguson-Smith, M.A., and Yang, F. (2008). Phylogenomics of the dog and fox family (Canidae, Carnivora) revealed by chromosome painting. Chromosom. Res. 16, 129–143. https://doi.org/10.1007/s10577-007-1203-5.
44. O'Connell, K.A., Yosufzai, Z.B., Campbell, R.A., Lobb, C.J., Engelken, H.T., Gorrell, L.M., Carlson, T.B., Catana, J.J., Mikdadi, D., Bonazzi, V.R., et al. (2023). Accelerating genomic workflows using NVIDIA Parabricks. BMC Bioinform. 24, 221. https://doi.org/10.1186/s12859-023-05292-2.
45. Pedersen, B.S., and Quinlan, A.R. (2017). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34, 867–868. https://doi.org/10.1093/bioinformatics/btx699.
46. Team, R.C. (2020). R: 2019. A Language and Environment for Statistical Computing version 3.
47. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
48. Pockrandt, C., Alzamel, M., Iliopoulos, C.S., and Reinert, K. (2020). GenMap: ultra-fast computation of genome mappability. Bioinformatics 36, 3687–3692. https://doi.org/10.1093/bioinformatics/btaa222.
49. Armstrong, E.E., and Campana, M.G. (2022). RatesTools: a Nextflow pipeline for detecting de novo germline mutations in pedigree sequence data. Bioinformatics 39, btac784. https://doi.org/10.1093/bioinformatics/btac784.
50. Lerat, E., Fablet, M., Modolo, L., Lopez-Maestre, H., and Vieira, C. (2017). TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. Nucleic Acids Res. 45, e17–e17. https://doi.org/10.1093/nar/gkw953.
51. Brůna, T., Li, H., Guhlin, J., Honsel, D., Herbold, S., Stanke, M., Nenasheva, N., Ebel, M.,

Gabriel, L., and Hoff, K.J. (2023). Galba: genome annotation with miniprot and AUGUSTUS. BMC Bioinform. 24, 327. https://doi.org/10.1186/s12859-023-05449-z.
52. Armstrong, E.E., Perry, B.W., Huang, Y., Garimella, K.V., Jansen, H.T., Robbins, C.T., Tucker, N.R., and Kelley, J.L. (2022). A beary good genome: Haplotype-resolved, chromosome-level assembly of the brown bear (Ursus arctos). Genome Biol. Evol. 14, evac125. https://doi.org/10.1093/gbe/evac125.
53. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. Science *376*, 44–53. https://doi.org/10.1126/science.abj6987.
54. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644. https://doi.org/10.1093/bioinformatics/btn013.
55. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
56. Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49, 303–309. https://doi.org/10.1038/ng.3748.
57. Stephan, S., and Ke, W. (2020). MSMC and MSMC2: The multiple sequentially Markovian coalescent. In Methods in Molecular Biology. (Springer US), pp. 147–166. https://doi.org/10.1007/978-1-0716-0199-0_7.
58. Liu, X., and Fu, Y.-X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21, 280. https://doi.org/10.1186/s13059-020-02196-9.
59. Koch, E.M., Schweizer, R.M., Schweizer, T.M., Stahler, D.R., Smith, D.W., Wayne, R.K., and Novembre, J. (2019). De novo mutation rate estimation in wolves of known pedigree. Mol. Biol. Evol. 36, 2536–2547. https://doi.org/10.1093/molbev/msz159.
60. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
61. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene set knowledge discovery with Enrichr. Curr. Protoc. 1, e90. https://doi.org/10.1002/cpz1.90.
62. Krassowski, M., Arts, M., Heeg, M., and Lagger, C. (2022). ComplexUpset.
63. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
64. Guarracino, A. (2024). paf2chain.
65. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730.