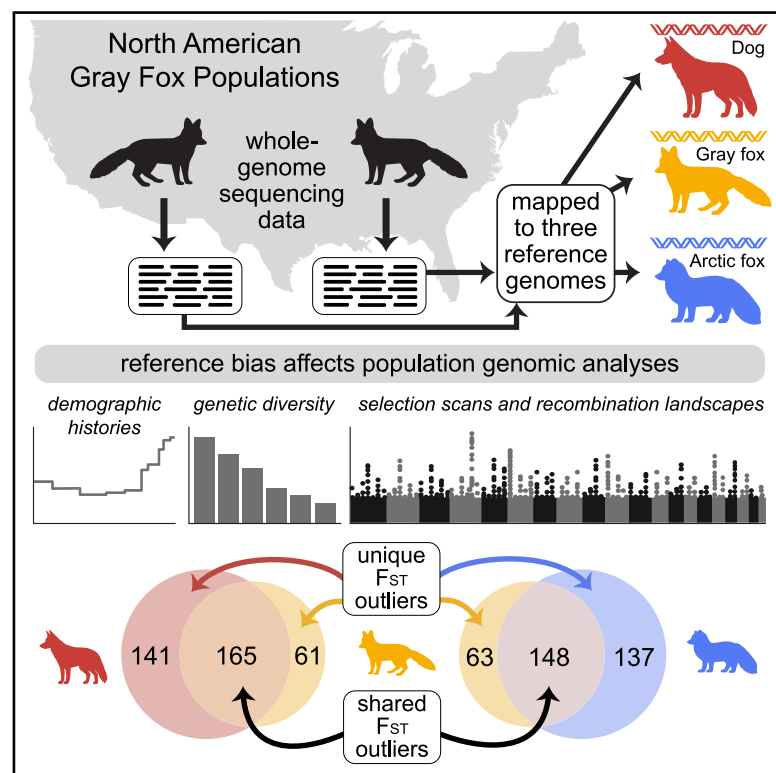


Reference genome choice compromises population genetic analyses

Graphical abstract



Authors

Maria Akopyan, Matthew Genchev, Ellie E. Armstrong, Jazlyn A. Mooney

Correspondence

jazlynmo@usc.edu

In brief

Using a species-specific reference genome significantly improves the accuracy of population genetic analyses, as exemplified by mapping gray fox sequence data to genomes of several canid species. Mapping to a conspecific genome provides more reliable estimates of genetic variation, diversity, and evolutionary history.

Highlights

- A species-specific reference genome improves read mapping and variant detection
- Reference bias underestimates genetic diversity and differentiation
- Divergent reference genomes distort demographic histories and recombination landscapes
- Unique F_{ST} outliers are detected across references, affecting functional interpretations

Article

Reference genome choice compromises population genetic analyses

Maria Akopyan,^{1,2} Matthew Genchev,¹ Ellie E. Armstrong,² and Jazlyn A. Mooney^{1,3,*}

¹Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA

²Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, Riverside, CA, USA

³Lead contact

*Correspondence: jazlynmo@usc.edu

<https://doi.org/10.1016/j.cell.2025.08.034>

SUMMARY

Characterizing genetic variation in natural populations is vital to evolutionary biology; however, many non-model species lack genomic resources. Here, we demonstrate that reference bias significantly affects population genomic analyses by mapping whole-genome sequence data from gray foxes (*Urocyon cinereoargenteus*) to a conspecific reference and two heterospecific canid genomes (dog and Arctic fox). Mapping to the conspecific genome improved read pairing by ~5% and detected 26%–32% more SNPs and 33%–35% more singletons. Nucleotide diversity estimates increased by over 30%, F_{ST} increased from 0.189 to 0.197, and effective population size estimates were 30%–60% higher with the conspecific reference. Recombination rates varied by up to 3-fold at chromosome ends with heterospecific references. Importantly, F_{ST} outlier detection differed markedly, with heterospecific genomes identifying twice as many unique outlier windows. These findings highlight the impact of reference genome choice and the importance of conspecific genomic resources for accurate evolutionary inference.

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 that 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. Reference genomes are essential for comparing individuals and populations, but they do not capture the full spectrum of genetic diversity. Population-level diversity is missed because a haploid reference represents variation from only a single individual at each site in the genome.

Reference bias arises because a single reference genome—usually a haploid sequence from one individual—is used as the coordinate system for mapping. This reference cannot fully represent the genetic diversity of an entire species. As a result,

sequencing reads that closely match the reference tend to map with higher quality and are retained, while dissimilar reads with lower quality scores often map poorly and are discarded. This can lead to mapping errors or missed variant calls, especially for true population-specific genetic variation. Thus, reference bias will result in missed or mis-called genotypes that will impact estimates of genetic diversity and downstream analyses, with the potential to distort conclusions about the population's evolutionary history. Reference bias, along with sequencing depth, has been shown to influence downstream genetic analyses. For example, in simulated low-coverage (2–4×) genomes, introducing a 2% reference divergence can cause estimates of θ to increase by 0.14- to 0.18-fold.⁴ Similar effects were noted on commonly used neutrality statistics: a small downward bias in Tajima's D , an upward bias in Fu and Li's D , and a strong downward bias in Fay and Wu's H .⁴ Importantly, these shifts were still observed even at moderate coverage (8×) and in empirical data.⁴

It is often the case that researchers will map to a reference genome from related species when a conspecific reference genome is unavailable. For example, in modern dogs and wolves, using a heterospecific (wolf) reference instead of a conspecific (dog) reference led to a 10% difference in heterozygosity estimates between individual dogs.⁵ In non-model species like big cats, using a conspecific reference can alter both heterozygosity (ranging from a 0.25-fold decrease to a 3.21-fold increase) and demographic history inference.⁶ In fishes,

reference bias has been associated with overestimated heterozygosity, with local reference genomes reducing reference/non-reference heterozygous calls by 0.13%–0.31% and non-reference/non-reference calls by 0.05%–0.10%.⁷ Studies have also shown that reference bias leads to inaccurate detection of structural variation⁸ and distorted phylogenies,⁹ potentially skewing conclusions about population history and genetic diversity in fish species. 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.¹⁰

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.¹¹ Previous gray fox genetic studies have relied on a domestic dog (*Canis lupus familiaris*) reference genome,¹² which poses challenges due to significant karyotypic differences between the two species.^{13,14} In fact, despite most canid studies' reliance on the domestic dog reference genome, the clade contains a number of large karyotypic rearrangements; for example, dogs have 38 pairs of autosomal chromosomes, gray foxes have 32, and the Arctic fox (*Vulpes lagopus*) has 23–24.

The domestic dog, gray fox, and Arctic fox represent distinct evolutionary lineages within *Canidae*.¹⁵ The gray fox represents the most basal lineage within the family, having diverged approximately 10 mya from all other living canids.¹⁵ The remaining canids are comprised of three major clades: the red fox clade, the South American clade, and the wolf-like clade. The Arctic fox belongs to the red fox clade, which diverged from the wolf-like clade approximately 7–8 mya. The domestic dog is part of the wolf-like clade and diverged most recently, having split from other wolf-like canids about 3–4 mya. There are also differences in the genome assembly sizes—the gray fox possesses a larger genome (2.66 Gb) than the dog (2.48 Gb) or the Arctic fox (2.35 Gb). Some of these differences may be due to the sex of the individual selected for assembly. Although GC content is relatively similar across all three species (~42%), repeat content is variable, with the gray fox exhibiting the highest percentage (38.23%), followed by the dog (33.67%) and the Arctic fox (31.29%). These genomic differences are likely to bias downstream haplotype-based analyses that rely on synteny and linkage disequilibrium (e.g., demographic inference and recombination mapping). Despite this, most previous genomic research within *Canidae*—including studies on common (gray fox), near-threatened (Channel Island fox),^{16,17} and endangered (Ethiopian wolf)¹⁸ species—has been conducted by mapping to the domestic dog genome, primarily because it was the only canid reference genome available with annotations and had both high continuity and contiguity.¹⁵

The recently published gray fox reference genome¹⁴ provides us with a high-quality conspecific reference for alignment, variant calling, and, ultimately, for quantifying reference bias. Here, we

re-analyze previously published WGS data from two populations of North American gray foxes.¹² We examine how inference of demographic histories and recombination landscapes differ when using the heterospecific domestic dog and 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, F_{ST} outlier scans for selection, and recombination rate inference.

RESULTS

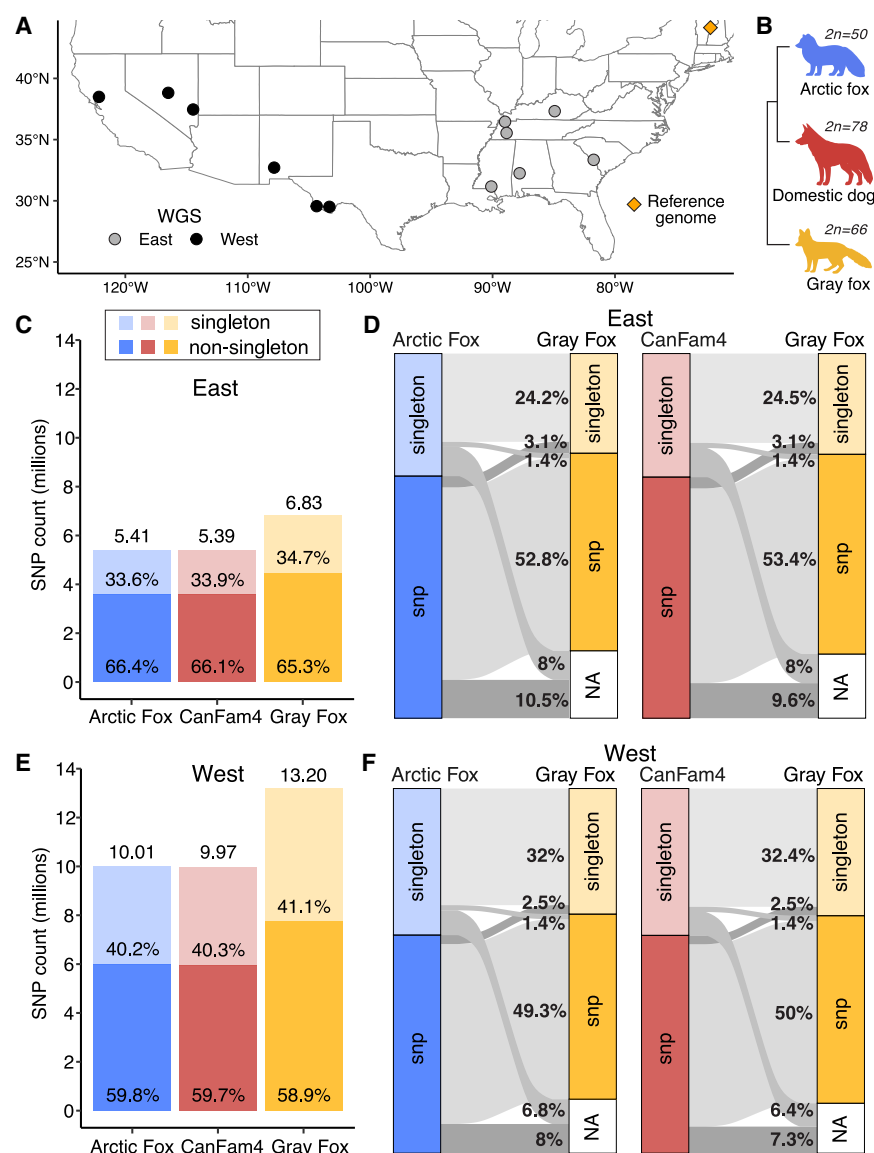
We analyzed whole-genome resequencing data from 12 gray foxes, including 6 from the eastern US and 6 from the western US populations (Figures 1A and S1). We examined reference bias by mapping these data to three reference genomes: the domestic dog genome CanFam4,¹⁹ the Arctic fox genome,²⁰ and the conspecific gray fox genome¹⁴ (Figure 1B). Gray fox reads showed significantly higher mapping success to the conspecific reference genome compared with heterospecific genomes (99.7% vs. ~99%; $\chi^2_{KW} = 33.19$, $p < 0.001$; Table 1), with genes in unmapped regions enriched for sensory perception and immunity functions. Additionally, proper read pairing was higher with the conspecific reference (94.7%) vs. heterospecific references (89.4%–90.3%), representing nearly 5% more properly paired reads ($\chi^2_{KW} = 40.14$, $p < 0.001$).

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 with the eastern population, with nearly double the number of SNPs identified in the west (Figure 1; Table 1). Notably, the gray fox reference genome consistently yielded more SNPs compared with the heterospecific references, with 26%–32% more variants identified in each population (Table 1; Figure 1).

We identified more singletons (i.e., rare variants where an allele is only found once in the population sample) in the east and west using the gray fox reference (Figures 1 and 2). Specifically, the conspecific reference detected 33% more singletons (2.4 M vs. 1.8 M) in the eastern population and 35% more singletons (5.4 M vs. 4 M) in the western population compared with the heterospecific references. In addition, mean SNP depth varied slightly across genomes, with gray fox showing higher values (mean: east; 8.5, west; 8.7) compared with Arctic fox and dog (both east: 8.3, west: 8.6). Missing data on a per-site basis were lowest with the gray fox genome (east and west: 0.02), whereas 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 1). Among these matching variants, non-singleton SNPs made up about 50%–53%, while the remainder were singletons. The western population had a higher proportion of matching singletons compared with that in the east (~32% vs. ~24%), whereas the eastern 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. Of these, about half mapped to invariant sites, while the other half did not map at all (Table S1).

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

We inferred demographic histories using *smc++*²¹ (sequential Markov coalescent) and observed distinct population size trajec-

tomies 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 (Figure 2). 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, although it should be noted that *smc++* typically has higher uncertainty for estimates in the recent past, <6,000 years ago.²¹ Meanwhile, the heterospecific genomes suggested a population decline, with N_e estimates dropping below 75,000.

To evaluate whether removing variants that failed to map between species improves demographic inference accuracy, we compared *smc++* trajectories before and after masking non-lifted-over SNPs in both heterospecific genomes and reciprocally in the gray fox genome. This approach made the demographic trajectories more similar across reference genomes (Figure S2). We also suspected that the total number of SNPs might affect demographic inference, so we downsampled the gray fox SNPs to random subsets matching the number of variants available when using lifted-over variants from other genomes. These downsampled trajectories were also similar to

Table 1. The total number of successfully mapped reads, number of SNPs, and heterozygosity estimates (mean \pm standard deviation) for each reference genome and population

Genome	Population	Reads mapped	Total SNPs	Heterozygosity (mean \pm SD)
Arctic fox	east	3,836,962,124	5,407,943	0.181 \pm 0.057
CanFam4	east	3,848,140,971	5,392,973	0.181 \pm 0.058
Gray fox	east	3,865,958,291	6,828,177	0.200 \pm 0.051
Arctic fox	west	1,032,573,693	10,012,122	0.163 \pm 0.049
CanFam4	west	1,035,433,543	9,969,686	0.163 \pm 0.049
Gray fox	west	1,040,317,760	13,199,882	0.174 \pm 0.047

The total number of reads (both mapped and unmapped) was 3,886,389,246 for the eastern population and 1,043,628,176 for the western population. Mapping success was significantly higher to the conspecific gray fox reference (99.7%) compared with heterospecific references (\sim 99%), with unmapped reads from heterospecific alignments enriched for genes involved in sensory perception and immunity, suggesting systematic exclusion of functionally important regions when using heterospecific references.

the lifted-over results, indicating that observed differences in demographic trajectories were driven not only by variant identity differences but also by the total number of informative variants available for analysis.

To evaluate how reference bias impacts demographic inference methods, we supplemented smc++ analysis (combining SFS with linkage disequilibrium in coalescent hidden Markov models [HMMs]) with MSMC2 (multiple sequentially Markovian coalescent, coalescent HMM only) and stairway plot 2 (SFS only) (Figure S2). Both MSMC2 and stairway plot 2 exhibited greater inconsistencies across reference genomes for the eastern population compared with the western population. MSMC2 inferred recent growth for the eastern population across all references, but heterospecific references exaggerated this growth, yielding eastern population sizes surpassing western ones—an outcome not observed with other methods and inconsistent with higher western diversity estimates. 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. Despite variation across reference genomes, the methodological approach had a stronger influence on demographic inference outcomes. The disparities between methods highlight the critical importance of selecting appropriate demographic reconstruction methods, as method-specific biases can overshadow reference genome variation and underscores the need for careful method selection.

Recombination rates vary across reference genomes and populations

We detected significant differences in the distributions of recombination rates, inferred using pyrho³ (see Table 2 for parameters), across reference genomes in both populations (Anderson-Darling k-sample test, $p < 0.001$; Figure 3). In the eastern population, recombination rates were lower when inferred using the Arctic fox genome compared with the gray fox genome. The average recombination rate was 0.448 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 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 with 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 east, where the heterospecific genomes led to both underestimation (Arctic fox) and overestimation (CanFam4) of recombination rates relative to the gray fox genome. In contrast, in the west, both heterospecific genomes consistently overestimated recombination rates compared with the gray fox genome, and the differences were less pronounced.

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. With the Arctic fox genome, 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 with 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 (Figure S3).

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 4). In the

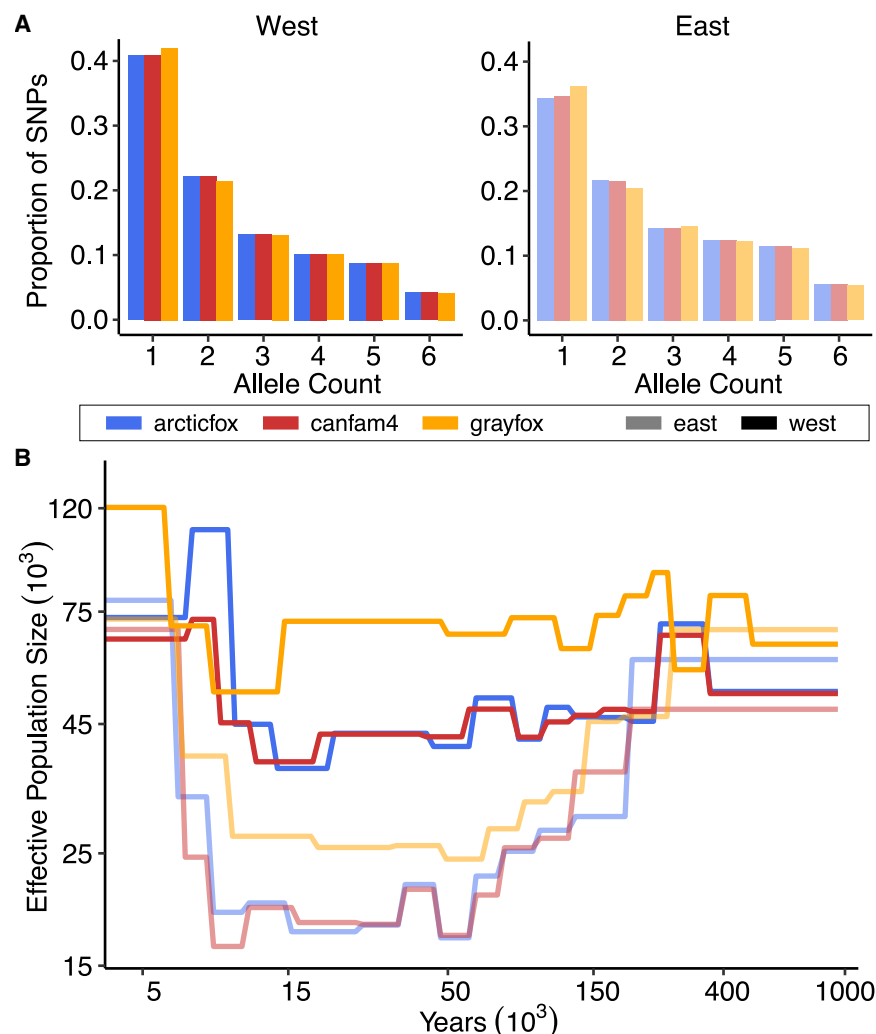


Figure 2. Reference bias influences the SFS and demographic trajectories

(A) SFS for eastern and western populations show more singletons using the conspecific gray fox genome.

(B) Effective population sizes (y axis) and years from present (x axis) inferred with *smc++* reveal discordant demographic histories of eastern (light) and western (dark) foxes resolved using the species-matched (gold) and heterospecific genomes (red and blue).

See also Figure S2.

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 ($\chi^2_{KW} = 558.27$, $p < 0.001$). Additionally, the correlation between nucleotide diversity and F_{ST} was more strongly negative in the eastern population ($\tau = -0.173$ to -0.196) compared with the western population ($\tau = -0.081$ to -0.118), with the strongest correlations observed using the gray fox reference in both populations.

In both populations, mean Tajima's D was lowest with the gray fox genome (-0.108 east, -0.356 west), compared with the Arctic fox (-0.056 east, -0.329 west) and CanFam4 (-0.063 east, -0.328 west) references ($\chi^2_{KW} = 195.3$ east, 144.4 west; $p < 0.001$). Pairwise comparisons confirmed significantly lower

eastern population, nucleotide diversity was lowest with the CanFam4 genome ($\pi = 0.000618$), followed closely by the Arctic fox ($\pi = 0.000625$), whereas the gray fox genome yielded a higher estimate of $\pi = 0.000812$ ($\chi^2_{KW} = 6,663.3$, $p < 0.001$). Similarly, in the western population, diversity was lowest using the heterospecific references (Arctic fox: $\pi = 0.00122$, CanFam4: $\pi = 0.00121$), whereas the gray fox genome produced a higher estimate of $\pi = 0.00164$ ($\chi^2_{KW} = 12,409$, $p < 0.001$).

The variation in estimates of nucleotide diversity was slightly more evident in the western population, where π was approximately 1.34 times higher using the gray fox genome compared with the heterospecific references, whereas in the eastern population π was 1.31 times higher. Although 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 that 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.

Tajima's D estimates with the gray fox genome ($p < 0.001$), whereas estimates did not differ between the two heterospecific references ($p = 0.17$ east, $p = 1.00$ west). This indicates an excess of low-frequency polymorphisms when mapping to the conspecific genome, consistent with the higher number of singletons detected in the SFS based on the gray fox reference.

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 with the conspecific reference (Figure 5). 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

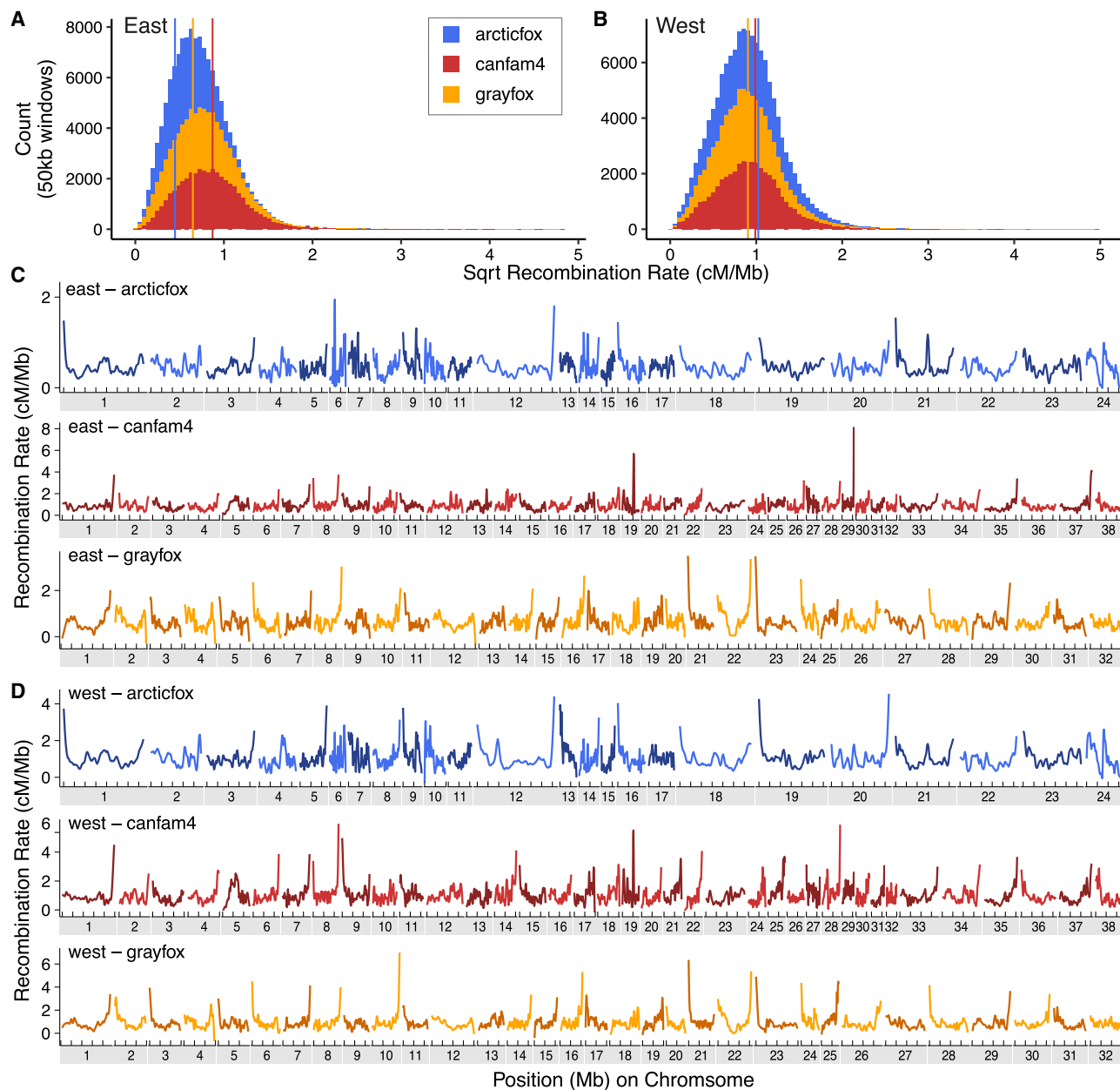


Figure 3. Recombination rate comparisons between reference genomes for gray fox populations

(A and B) Histograms of the square root of recombination rates (cM/Mb) across 50-kb windows in (A) eastern and (B) western populations. Colors represent the three reference genomes used for recombination rate estimates, with vertical lines representing mean values.

(C and D) 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.

See also Figure S3.

the gray fox reference (Figure 5A). 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 5B). 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.

Between the Arctic fox and Canfam4, we identified 191 shared outlier windows, 158 windows unique to the Arctic fox reference, and 42 windows unique to the Canfam4 reference, with 35,133 windows showing no elevated F_{ST} values in either reference. Thus, the Arctic fox and Canfam4 references shared slightly more outliers with each other than either shared with the gray fox (148–165), while the number of unique outliers for each

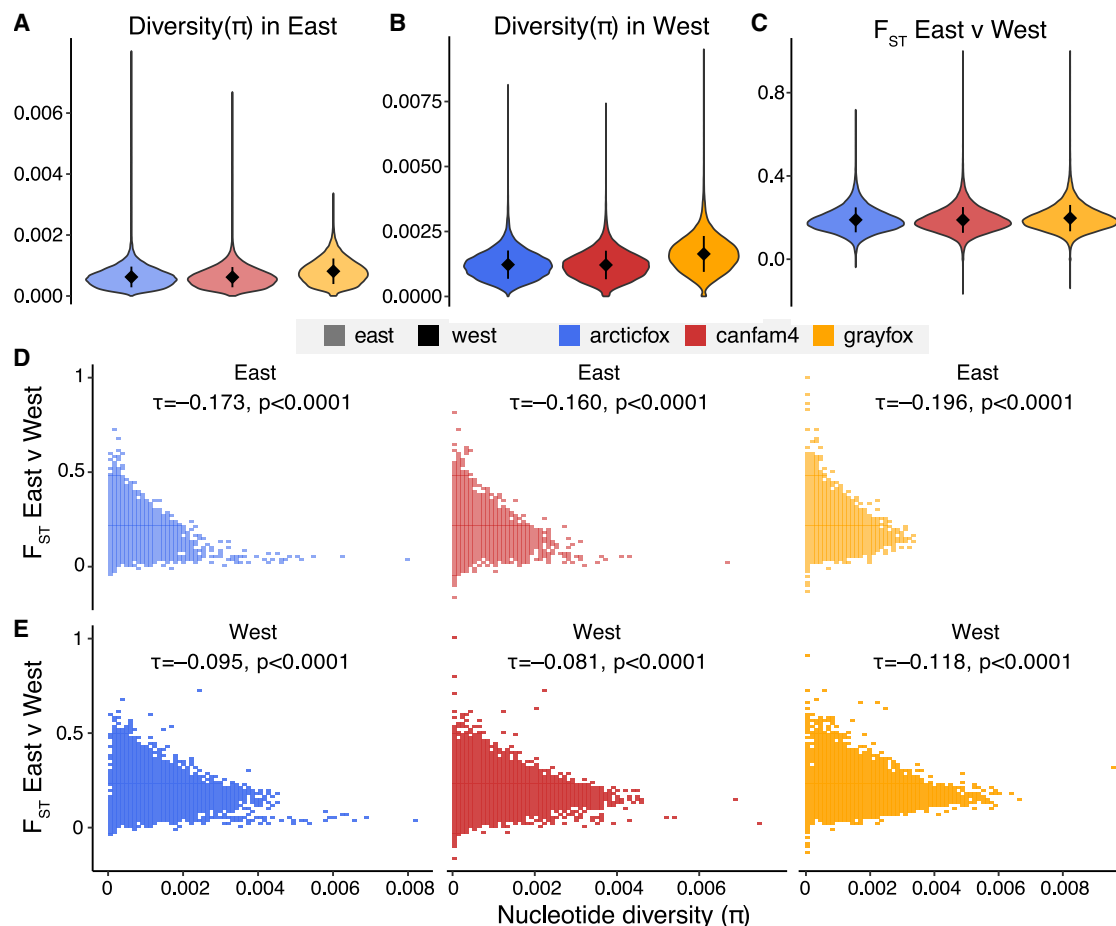


Figure 4. Estimates of diversity and differentiation are higher with the gray fox reference in eastern and western gray fox populations

(A–C) 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.

(D and E) The correlation between diversity and differentiation across genomic windows for the eastern (D) and western (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.

See also Figure S4.

heterospecific reference remained comparable to their gray fox comparisons.

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

The functional enrichment analysis of F_{ST} outliers revealed both distinct and overlapping Gene Ontology (GO) biological process terms across reference genomes (Figure 4C). The gray fox had the highest number of unique terms with 74, followed by Canfam4 with 71, and the 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 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.”

GO enrichment analysis of genes located exclusively in Arctic fox reference-specific outlier windows revealed significant enrichment for terms related to muscle development and metabolism. The top biological process terms included “muscle

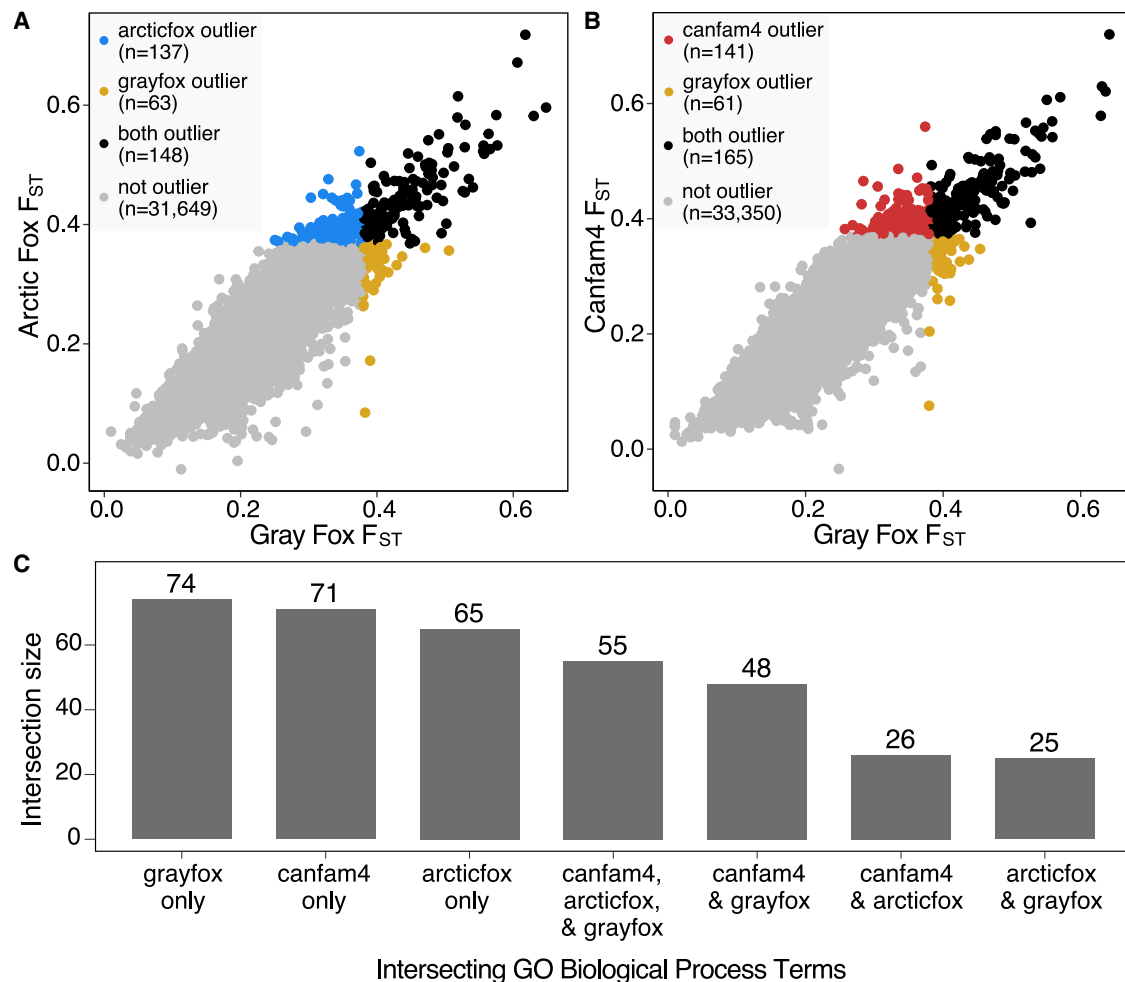


Figure 5. Reference-specific patterns revealed by F_{ST} outlier and functional enrichment analyses

(A) Scatterplot comparing F_{ST} values between the gray fox and arctic fox references, with outlier windows unique to the arctic fox (blue) and gray fox (gold) highlighted.

(B) Scatterplot comparing F_{ST} values between the 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 (y axis) of GO terms shared across reference genomes (x axis).

See also Figure S4.

organ development" (GO:0007517), "positive regulation of muscle hypertrophy" (GO:0014742), and "positive regulation of glycolytic process" (GO:0045821). In contrast, genes in Canfam4 reference-specific outlier windows showed markedly different functional enrichment patterns, predominantly associated with chromosomal organization and cellular differentiation. The most significant terms included "positive regulation of chromosome organization" (GO:2001252), "cell differentiation in spinal cord" (GO:0021515), and "regulation of chromosome separation" (GO:1905818). These distinct functional enrichment patterns demonstrate how reference genome choice can lead to substantially different biological interpretations. The Arctic fox reference-specific outliers suggest selection on genes involved in muscle development and energy metabolism, which could be interpreted as adaptations related to locomotion, hunt-

ing behavior, or thermal regulation. Conversely, the Canfam4 reference-specific outliers point toward selection on chromosomal organization and cellular differentiation pathways, potentially indicating differences in developmental processes or cell cycle regulation.

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 chromosome-level assemblies,²² forcing researchers to use genomes from closely related species,^{23,24} which can compound reference bias. Reference bias has been previously studied across the tree of

life,^{5,6,10,25–28} both within species^{26,27} and between species.⁶ 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 the 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 is 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, including retaining only reads with a certain mapping quality (MAPQ) score, retaining proper pairs, or only retaining the primary alignment.²⁹ Likely, the number of properly paired reads decreases when the gray fox is mapped to the heterospecific genomes because of structural variation between the genomes, some of which may be smaller insertions or deletions (indels), and some of which may reflect the large chromosomal rearrangements between the species. Following common practice, we retained improperly paired reads, which may explain downstream variant detection effects, though selective filtering could potentially mitigate this.

To exemplify the inconsistency of pipelines and filters applied, we screened five papers that 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,^{12,16,18,30} one paper did remove improperly paired reads,¹⁷ and the 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 Burrows-Wheeler Aligner (BWA)³¹ and not as a pair (i.e., the MAPQ is ambiguous to the read being properly paired). The genome analysis toolkit (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 variant confidence (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 more SNPs and a more reliable SFS compared with the heterospecific genomes. Heterospecific references underestimated the presence of low-frequency alleles, which can have profound implications for downstream analyses. For example, underestimating rare variants can lead to failure to detect recent population expansion in demographic inference, as the characteristic excess of rare alleles that signals expansion would be artificially reduced. Additionally, an underestimation of genetic load can occur when assessing deleterious mutations, as harmful variants tend to be maintained at low frequencies due to purifying selec-

tion. Furthermore, recombination rate estimates in population-scale maps can become distorted due to missing variants, as these estimates rely on accurate measurement of linkage disequilibrium patterns between neighboring loci.

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 1). Of these, ~5% were misclassified (singleton vs. non-singleton) and ~15% were unrecognized as variants, with half mapping to monomorphic sites and half failing to map (Table S1). 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 an overall larger and more 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 with the eastern population, suggesting differing evolutionary pressures and histories in these lineages. Our results differ from previous studies¹² of these populations, which mapped to CanFam3.1 and used the pairwise sequentially Markovian coalescent (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 eastern and western 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 time points, though the magnitude of our diversity estimates is much lower than previous estimates.¹² 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.¹²

We used multiple methods for demographic inference to determine whether coalescent or SFS-based approaches were more robust to reference bias. All methods showed that heterospecific genomes consistently underestimated effective population sizes compared with the conspecific reference. The conspecific reference produced estimates that align more closely with expectations based on species biology and similar studies in other mammals.³² Importantly, heterospecific genomes

produced demographic trajectories inconsistent with the conspecific reference throughout most of the inferred history, regardless of population or inference method (Figure S2). This inconsistency is particularly worrying in the case of species threatened with extinction because approximately 99% of these species do not have a reference genome³³ and researchers will ultimately be forced to map to a heterospecific reference. When researchers must use heterospecific references, several strategies can mitigate reference bias. For instance, using end-to-end alignment methods that require the entire read to align reduces bias at indels, compared with local aligners that allow parts of reads that do not align to be ignored (i.e., soft clipped).³⁴ Using local realignment methods that ensure consistent gap placements across all reads covering the same region, and comparing results across different filtering thresholds can also help reduce bias.³⁴ Additionally, constructing consensus references or using multiple population reference genomes in a “reference flow” approach effectively reduces bias while requiring fewer computational resources than graph-based methods.³⁵

Because pedigree-based recombination estimates are difficult to obtain and have low resolution, researchers typically use population-based methods like *pyrho*³ to infer recombination maps that account for demographic histories. 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 toward the end of the chromosome, has been previously observed in dogs³⁶ and is particularly interesting given that the family *Canidae* has a pseudogenized copy of positive-regulatory-domain zinc-finger protein 9 (PRDM9).^{14,37} PRDM9 is known to initiate meiotic recombination by specifying the locations of double-strand breaks and creating recombination hotspots. Previous work highlighted how its pseudogenization in *Canidae* contributes to observed recombination patterns in dogs, which can be used to identify the directionality of chromosomes.³⁶ Chromosome directionality has yet to be established for the 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 with the gray fox genome, whereas 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 with 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.¹⁴ 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 S2). 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 S3). 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*³ 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 1) 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 with 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.

Table 2. Pyrro hyperparameter settings for each reference genome and both populations

Reference genome	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

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 with the conspecific reference. This discrepancy could be due to alignment issues, where regions not well conserved between species may erroneously appear as highly differentiated regions. 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 the F_{ST} outliers, genes, and biological pathways that may be under selection. Using a conspecific reference would certainly provide more accurate insights into the biological processes influencing genetic differentiation within and between species.

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.^{30,39} Furthermore, genomic data are increasingly being used to inform conservation management plans. Measures such as adaptive capacity⁴⁰ and differential adaptations between populations⁴¹ are being used to make recommendations regarding translocations and rewilding. Our selection scans suggest that these inferences are 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 heter-

ospecific 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.

Limitations of the study

Our work provides important insights into reference bias but has limitations. First, the gray fox reference genome comes from an east coast individual, potentially introducing bias that future work could mitigate computationally or by generating a west coast reference. Second, we quantified reference bias in *Canidae*, a mammalian clade with notable karyotype shifts. Despite these shifts, *Canidae* species are relatively young, likely moderating reference bias compared with more divergent taxa. More divergent non-mammalian taxa are likely to experience stronger reference bias. Because non-mammalian species comprise most of the 99% of species without reference genomes, this highlights the broader importance of addressing reference bias. However, karyotype differences occur even between closely related species, making *Canidae* useful for studying the interplay between karyotype shifts and reference bias. Additionally, our mapping and filtering parameters, while representing common practices, may influence the magnitude of reference bias observed. Further analysis is needed to examine how these choices interact with reference bias and impact other genomic analyses, including inference of introgression patterns and linkage disequilibrium, which were beyond the scope of this study. This study does not address reference bias in ancient DNA. Previous research has shown that, in addition to DNA quality, reference bias is influenced by factors such as sequence length, genetic divergence,^{42,43} the alignment tool used,⁴² and map quality filtering.^{42,44} Genetic summary statistics require careful interpretation in ancient DNA studies, where both human and non-human extinct lineages will be mapped to divergent interspecific references. The additional temporal component of ancient DNA will likely amplify the biases observed here.

Conclusions

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. Considering these biases, our work demonstrates that results generated from mapping to a highly divergent reference genome should be interpreted with caution.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Jazlyn A. Mooney (jazlynmo@usc.edu).

Materials availability

This study did not generate new reagents.

Data and code availability

- All original code has been deposited at GitHub and is publicly available as of the date of publication. Code for data filtering can be found at https://github.com/ellieearmstrong/Gray_Fox_2023/tree/main/filtering, and code for analyses can be found at <https://github.com/makopyan/fox/>.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

We want to thank Doc Edge, Joshua Schraiber, and Jeffrey Spence for insightful and helpful conversations during this work. J.A.M. and M.A. were 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.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Whole genome sequence data and reference genomes
 - Sample selection
- **METHOD DETAILS**
 - Mapping and variant calling
 - Variant filtering
 - Genome annotations
 - Functional enrichment analysis of unmapped reads
 - Computing genetic diversity (π) and differentiation (F_{ST})
 - Quantifying genetic variation
 - Demographic inference
 - Recombination rate estimation
 - Identifying F_{ST} outliers and conducting functional enrichment analyses
 - Converting coordinates between reference genomes

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2025.08.034>.

Received: November 26, 2024

Revised: April 24, 2025

Accepted: August 28, 2025

REFERENCES

- 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, 117. <https://doi.org/10.1186/s13059-019-1720-5>.
- Salmona, J., Heller, R., Lascoux, M., and Shafer, A. (2017). Inferring Demographic History Using Genomic Data. In *Population Genomics*, O.P. Rajora, ed. (Springer), pp. 511–537. https://doi.org/10.1007/13836_2017_1.
- 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>.
- Nevado, B., Ramos-Onsins, S.E., and Perez-Enciso, M. (2014). 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>.
- Gopalakrishnan, S., Samaniego Castruita, J.A., Sinding, M.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 Genomics* 18, 495. <https://doi.org/10.1186/s12864-017-3883-3>.
- 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/s12915-019-0734-5>.
- 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>.
- 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. Preprint at bioRxiv. <https://doi.org/10.1101/2024.06.28.599671>.
- Rick, J.A., Brock, C.D., Lewanski, A.L., Golcher-Benavides, J., and Wagner, C.E. (2024). Reference genome choice and filtering thresholds jointly influence phylogenomic analyses. *Syst. Biol.* 73, 76–101. <https://doi.org/10.1093/sysbio/syad065>.
- 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>.
- 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>.
- 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>.
- Wayne, R.K., Nash, W.G., and O'Brien, S.J. (1987). Chromosomal evolution of the Canidae. I. Species with high diploid numbers. *Cytogenet. Cell Genet.* 44, 123–133. <https://doi.org/10.1159/000132356>.
- 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 (Bethesda)* 14, jkae034. <https://doi.org/10.1093/g3journal/jkae034>.
- 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>.

16. Robinson, J.A., Ortega-Del Vecchyo, D.O.-D., Fan, Z., Kim, B.Y., von Holdt, 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>.
17. 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>.
18. Mooney, J.A., Marsden, C.D., Yohannes, A., Wayne, R.K., and Lohmueller, K.E. (2023). Long-term small population size, deleterious variation, and altitude adaptation in the Ethiopian wolf, a severely endangered canid. *Mol. Biol. Evol.* 40, msac277. <https://doi.org/10.1093/molbev/msac277>.
19. Wang, C., Wallerman, O., Arendt, M.-L., Sundström, E., Karlsson, Å., Norddin, 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>.
20. 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/10.1111/1755-0998.13397>.
21. 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>.
22. Lewin, H.A., Richards, S., Lieberman 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. USA* 119, e2115635118. <https://doi.org/10.1073/pnas.2115635118>.
23. Galla, S.J., Forsdick, N.J., Brown, L., Hoepfner, M.P., 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>.
24. 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>.
25. 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>.
26. 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>.
27. Miga, K.H., and Wang, T. (2021). The Need for a Human Pangenome Reference Sequence. *Annu. Rev. Genomics Hum. Genet.* 22, 81–102. <https://doi.org/10.1146/annurev-genom-120120-081921>.
28. 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>.
29. 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>.
30. Chavez, D.E., Gronau, I., Hains, T., Dikow, R.B., Frandsen, P.B., Figueiró, H.V., Garcez, F.S., Tchaicka, L., de Paula, R.C., 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. USA* 119, e2205986119. <https://doi.org/10.1073/pnas.2205986119>.
31. 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>.
32. 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. <https://doi.org/10.1126/science.abn5856>.
33. Paez, S., Kraus, R.H.S., Shapiro, B., Gilbert, M.T.P., Jarvis, E.D., Vertebrate Genomes Project; Conservation Group, 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>.
34. Lin, M.-J., Iyer, S., Chen, N.-C., and Langmead, B. (2024). Measuring, visualizing, and diagnosing reference bias with biastools. *Genome Biol.* 25, 101. <https://doi.org/10.1186/s13059-024-03240-8>.
35. Chen, N.-C., Solomon, B., Mun, T., Iyer, S., and Langmead, B. (2021). Reference flow: reducing reference bias using multiple population genomes. *Genome Biol.* 22, 8. <https://doi.org/10.1186/s13059-020-02229-3>.
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 (Bethesda)* 6, 3517–3524. <https://doi.org/10.1534/g3.116.034678>.
37. Axelsson, E., Webster, M.T., Ratnakumar, A., LUPA Consortium, 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, g1aa134. <https://doi.org/10.1093/giga-science/g1aa134>.
39. 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>.
40. 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>.
41. 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. Change Resp.* 2, 1. <https://doi.org/10.1186/s40665-014-0009-x>.
42. Dolenz, S., van der Valk, T., Jin, C., Oppenheimer, J., Sharif, M.B., Orlando, L., Shapiro, B., Dalén, L., and Heintzman, P.D. (2024). Unravelling reference bias in ancient DNA datasets. *Bioinformatics* 40, btae436. <https://doi.org/10.1093/bioinformatics/btae436>.
43. Swiel, Y., Kelso, J., and Peyrégne, S. (2025). Resolving the source of branch length variation in the Y chromosome phylogeny. *Genome Biol.* 26, 4. <https://doi.org/10.1186/s13059-024-03468-4>.
44. 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>.
45. 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 Bioinformatics* 24, 221. <https://doi.org/10.1186/s12859-023-05292-2>.

46. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34, 867–868. <https://doi.org/10.1093/bioinformatics/btx699>.
47. R Core Team (2020). R: A Language and Environment for Statistical Computing Version 3 (R Foundation for Statistical Computing).
48. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686. <https://doi.org/10.21105/joss.01686>.
49. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernyt-sky, 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>.
50. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
51. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.
52. Pockrandt, C., Alzamel, M., Iliopoulos, C.S., and Reinert, K. (2020). Gen-Map: ultra-fast computation of genome mappability. *Bioinformatics* 36, 3687–3692. <https://doi.org/10.1093/bioinformatics/btaa222>.
53. 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. <https://doi.org/10.1093/nar/gkw953>.
54. Brūna, T., Li, H., Guhlén, J., Honsel, D., Herbold, S., Stanke, M., Nena-sheva, 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>.
55. 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>.
56. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>.
57. 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>.
58. Schiffels, S., and Wang, K. (2020). MSMC and MSMC2: The multiple sequentially Markovian coalescent. In *Methods in Molecular Biology*, 2090 (Springer), pp. 147–166. https://doi.org/10.1007/978-1-0716-0199-0_7.
59. 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>.
60. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
61. 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>.
62. 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>.
63. Krassowski, M., Arts, M., Heeg, M., and Lagger, C. (2022). ComplexUpset. Zenodo. <https://zenodo.org/records/7314197>.
64. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowski, 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>.
65. Guarracino, A. (2024). paf2chain. <https://github.com/AndreaGuarracino/paf2chain>.
66. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>.
67. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed data-sets. *Bioinformatics* 26, 2204–2207. <https://doi.org/10.1093/bioinformatics/btq351>.
68. Armstrong, E.E., and Campana, M.G. (2023). RatesTools: a Nextflow pipeline for detecting de novo germline mutations in pedigree sequence data. *Bioinformatics* 39, btac784. <https://doi.org/10.1093/bioinformatics/btac784>.
69. 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>.
70. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizikadze, 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>.
71. Meadows, J.R.S., Kidd, J.M., Wang, G.-D., Parker, H.G., Schall, P.Z., Bianchi, M., Christmas, M.J., Bougiouri, K., Buckley, R.M., Hitte, C., et al. (2023). Genome sequencing of 2000 canids by the Dog10K consortium advances the understanding of demography, genome function and architecture. *Genome Biol.* 24, 187. <https://doi.org/10.1186/s13059-023-03023-7>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
NVIDIA Parabricks v4.1.1.144	O'Connell et al. ⁴⁵	https://www.nvidia.com/en-us/clara/genomics/
Mosdepth v0.3.3	Pedersen and Quinlan ⁴⁶	https://github.com/brentp/mosdepth ; RRID: SCR_018929
SAMtools v1.16.1	Li and Durbin ³¹	http://www.htslib.org/ ; RRID: SCR_002105
R v4.2.1/v4.3.2	R Core Team ⁴⁷	https://www.r-project.org/ ; RRID: SCR_001905
Tidyverse	Wickham et al. ⁴⁸	https://www.tidyverse.org/ ; RRID: SCR_019186
GATK v4.1.4.1	McKenna et al. ⁴⁹	https://gatk.broadinstitute.org/ ; RRID: SCR_001876
pysam v0.22.1	Li and Durbin ³¹	https://github.com/pysam-developers/pysam ; RRID: SCR_021017
bedtools v2.31.1	Quinlan and Hall ⁵⁰	http://bedtools.readthedocs.io/ ; RRID: SCR_006646
BCFtools v1.16	Danecek et al. ⁵¹	http://samtools.github.io/bcftools/ ; RRID: SCR_005227
GenMap v1.3.0	Pockrandt et al. ⁵²	https://github.com/cpockrandt/genmap
TETools v1.7	Lerat et al. ⁵³	https://github.com/l-modolo/TEtools
GALBA v1.0.9	Brúna et al. ⁵⁴	https://github.com/Gaius-Augustus/GALBA
AUGUSTUS v3.3.2	Stanke et al. ⁵⁵	https://github.com/Gaius-Augustus/Augustus ; RRID: SCR_008417
SNPRelate v1.36.1	Zheng et al. ⁵⁶	https://github.com/zhengxwen/SNPRelate ; RRID: SCR_022719
VCFtools v1.14	Danecek et al. ⁵⁷	https://vcftools.github.io/ ; RRID: 001235
smc++ v1.15.4	Terhorst et al. ²¹	https://github.com/popgenmethods/smcpp
MSMC2 v2.1.4	Schiffels and Wang ⁵⁸	https://github.com/stschiff/msmc2 ; RRID: SCR_023677
MSMC-tools	Schiffels and Wang ⁵⁸	https://github.com/stschiff/msmc-tools
Stairway Plot 2 v2.1.2	Liu and Fu ⁵⁹	https://github.com/xiaoming-liu/stairway-plot-v2
pyrho v0.1.0	Spence and Song ³	https://github.com/popgenmethods/pyrho
minimap2 v2.14	Li ⁶⁰	https://github.com/lh3/minimap2 ; RRID: SCR_018550
Crossmap v0.7.0	Zhao et al. ⁶¹	https://crossmap.sourceforge.net/ ; RRID: SCR_001173
GenomicRanges v1.54.1	Lawrence et al. ⁶²	https://bioconductor.org/packages/GenomicRanges ; RRID: SCR_017051
ComplexUpset v1.3.3	Krassowski et al. ⁶³	https://github.com/krassowski/complex-upset ; RRID: SCR_022752
Enrichr	Xie et al. ⁶⁴	https://maayanlab.cloud/Enrichr/ ; RRID: SCR_001575
paf2chain	Guarracino ⁶⁵	https://github.com/AndreaGuarracino/paf2chain

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SnEff	Cingolani et al. ⁶⁶	http://pcingola.github.io/SnpEff/ ; RRID: SCR_005191
bigWigAverageOverBed	Kent et al. ⁶⁷	https://genome.ucsc.edu/goldenPath/help/bigWig.html
filterGM.rb	Armstrong and Campana ⁶⁸	https://github.com/elliiearmstrong/Gray_Fox_2023
RM2bed.rb	Armstrong and Campana ⁶⁸	https://github.com/elliiearmstrong/Gray_Fox_2023
simplify_bed.rb	Armstrong and Campana ⁶⁸	https://github.com/elliiearmstrong/Gray_Fox_2023

Deposited data

Gray fox whole-genome data	Preckler-Quisquater et al. ¹²	GenBank: PRJNA966176
Gray fox reference genome	Armstrong et al. ¹⁴	GenBank: GCA_032313775.1
Arctic fox reference genome	Peng et al. ²⁰	GenBank: GCF_018345385.1
Domestic dog reference genome (CanFam4)	Wang et al. ¹⁹	GenBank: GCF_011100685.1
Brown bear genome	Armstrong et al. ⁶⁹	GenBank: GCF_023065955.2
Human genome assembly	Nurk et al. ⁷⁰	GenBank: GCF_009914755.1
PhyloP scores for CanFam4	Meadows et al. ⁷¹	Zenodo: 8084059

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

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 Koguchi et al.,¹ in which the data were mapped to the domestic dog reference genome CanFam3.1.² We re-mapped the data to three reference genomes: (1) Arctic fox,³ (2) a more contiguous updated version of the dog genome CanFam4,⁴ and (3) our recently published gray fox reference genome,⁵ based on an individual sampled in Vermont (Figure 1A). We identified syntenic and non-syntenic chromosomes between species using information from Armstrong et al.⁶ and through alignments generated in Gopalakrishnan et al.⁵ All putatively syntenic chromosomes from Armstrong et al.⁶ were then manually interrogated using the genomic alignments from Gopalakrishnan et al.⁵ 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.

Sample selection

We performed a principal component analysis (PCA) on SNP genotype data including all 41 individuals using the SNPRelate v1.36.1¹³ package in R v4.3.2.⁹ We first converted the VCF file (see method details below) 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.

METHOD DETAILS

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 v4.1.1.1⁷ `fq2bam` with the provided reference and default parameters. We estimated depth and other mapping statistics using `Mosdepth` v0.3.3⁸ 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 v4.2.1.⁹ 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 v4.1.4.1¹⁰ `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 GATK `GenotypeGVCFs` with the `-all-sites` flag to produce final gVCF files. Files

were combined using BCFtools v1.16¹¹ *concat* to create genome-wide variant calls. Read mapping counts were generated using pysam v0.22.1¹² and custom scripts.

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/elliiearmstrong/Gray_Fox_2023/tree/main/filtering.

Genome annotations

We produced mappability scores for each genome using GenMap v1.3.0.¹⁴ 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 Lindblad-Toh et al.¹⁵ to generate a file of sites with a mappability score < 1.

We also generated repetitive element annotations for each genome. Briefly, we ran TETools v1.7¹⁶ on each genome assembly as in Gopalakrishnan et al.⁵ 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.¹⁵

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 v1.0.9¹⁷ 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 (GenBank: GCF_018345385.1³), domestic dog (GenBank: GCF_011100685.1⁴), brown bear (GenBank: GCF_023065955.2¹⁸), and the most recent (at the time) human genome assembly (GenBank: GCF_009914755.1¹⁹). These files were provided as protein evidence to GALBA using the –prot_seq flag and run using AUGUSTUS v3.3.2²⁰ which was run with default parameters.

Functional enrichment analysis of unmapped reads

To determine whether the reads that did not map to heterospecific reference genomes but mapped to the conspecific genome were enriched for functional regions, we used the read tags (unique read identifiers) of the unmapped reads from Arctic fox and CanFam4 genomes to obtain the positions where they mapped in the gray fox genome. We used bedtools v2.31.1²¹ to intersect those positions with the gray fox genome annotation, then conducted a functional enrichment analysis for gene ontology biological processes using Enrichr²² to identify biological processes significantly overrepresented in those regions.

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 VCFtools²³ 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. Additionally, to compare the number of segregating sites per site with nucleotide diversity, we calculated Tajima’s D using VCFtools v1.14 in non-overlapping 50 kb windows based on variant sites for each population and reference genome.

Quantifying genetic variation

We used VCFtools²³ 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 package²⁴ in R v. 4.3.2.⁹ 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++`²⁵ v1.15.4, `MSMC2`²⁶ v2.1.4, and `Stairway Plot 2` v2.1.2.²⁷ `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 `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 $4.5\text{e-}9$ ²⁸ 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 years²⁹ 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 $4.5\text{e-}9$,²⁸ a generation time of two years,²⁹ 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, including a negative mask to remove regions with low mappability scores. We estimated population size histories by specifying pair indices (-l) for each individual, to avoid pairs of haplotypes from different individuals, as recommended for unphased genomes.

To evaluate whether demographic reconstructions improve when using SNPs that successfully lifted over from the heterospecific genomes to the gray fox genome, we re-ran `smc++` after masking variants in the Arctic fox and dog genomes that failed to liftover. We then compared these masked demographic trajectories to the original reconstructions.

Recombination rate estimation

Recombination rates were estimated using `pyrho` v0.1.0,³⁰ which takes in the demography from `smc++` and linkage information from unphased data to estimate population-specific fine-scale recombination rates per generation.³⁰ 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 ($4.5\text{e-}9$ ²⁸). 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 2), 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 $1\text{e}8$ (100 cM per expected crossover multiplied by $1\text{e}6$ bases per mb). To compare recombination hotspots across references, we converted window positions between each heterospecific reference and the gray fox reference (described below). Hotspots were defined as windows with recombination rates exceeding two standard deviations above the mean for each reference genome and categorized based on whether they were shared across references or unique to a specific reference.

Identifying F_{ST} outliers and conducting 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). We also extended this analysis to compare F_{ST} outlier detection between the two heterospecific genomes. 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 GenomicRanges³¹ v1.54.1 package in R v. 4.3.2⁹ 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 Enrichr²² to identify processes significantly overrepresented in the foreground gene set compared to the background set. We used the ComplexUpset³² v1.3.3 R package to create upset plots showing the number of identified genes from each reference and their intersection. To investigate the impact of reference genome choice on F_{ST} outlier detection, we examined the specific genes located within unique outlier windows for each heterospecific reference genome. This analysis aimed to determine whether biologically plausible interpretations could be derived from potentially reference-biased regions.

To assess the impact of reference choice on outlier classification, we examined multiple genomic features within matched outlier windows, including evolutionary constraint scores, gene annotations, and the distribution of synonymous and non-synonymous mutations. We obtained PhyloP constraint scores for CanFam4 from the International Dog10K project.³³ Using bigWigAverageOverBed,³⁴ we calculated the mean PhyloP score for each 50kb window, matching the F_{ST} windows that we lifted over to align across reference genomes using the GenomicRanges package in R. Additionally, we evaluated the genic and intergenic composition of outlier windows based on annotations from each reference genome. Next, we used SnpEff³⁵ to assess the functional impact of variants within outlier windows by building custom annotation databases for each reference genome using its corresponding gene annotations. We annotated variants in the VCF files separately for each reference, classifying mutations as synonymous or non-synonymous based on their predicted coding effects.

Converting coordinates between reference genomes

We performed whole genome alignments between the two heterospecific genomes and the gray fox genome using minimap2³⁶ v2.14 with parameters `-cs -x asm20`. We converted the pairwise alignment format computed by minimap2 to a chain file using paf2chain,³⁷ then used the Crossmap³⁸ 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.

Supplemental figures

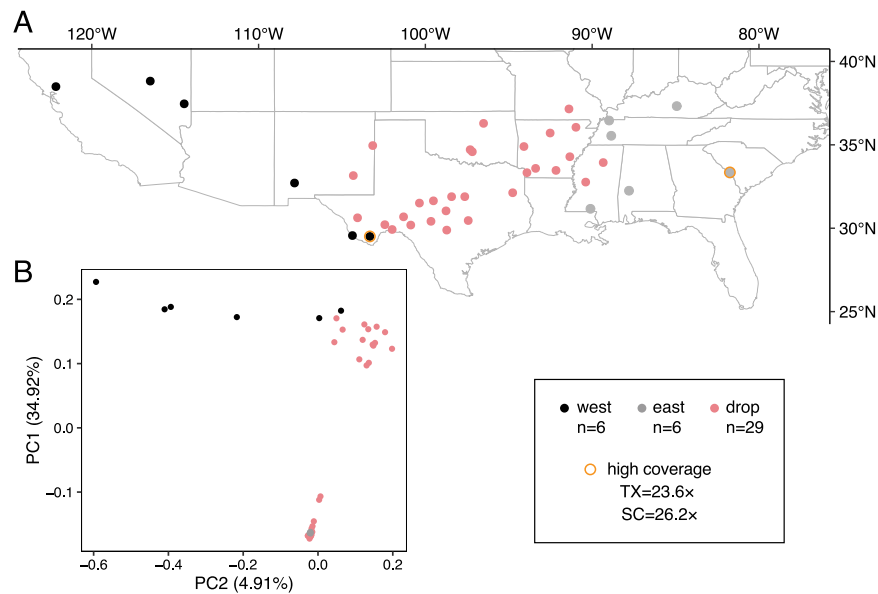


Figure S1. Sampling locations of genomes, related to Figure 1

(A) Sampling localities of WGS data for $n = 41$ gray foxes obtained from Kosugi et al.,¹ 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.

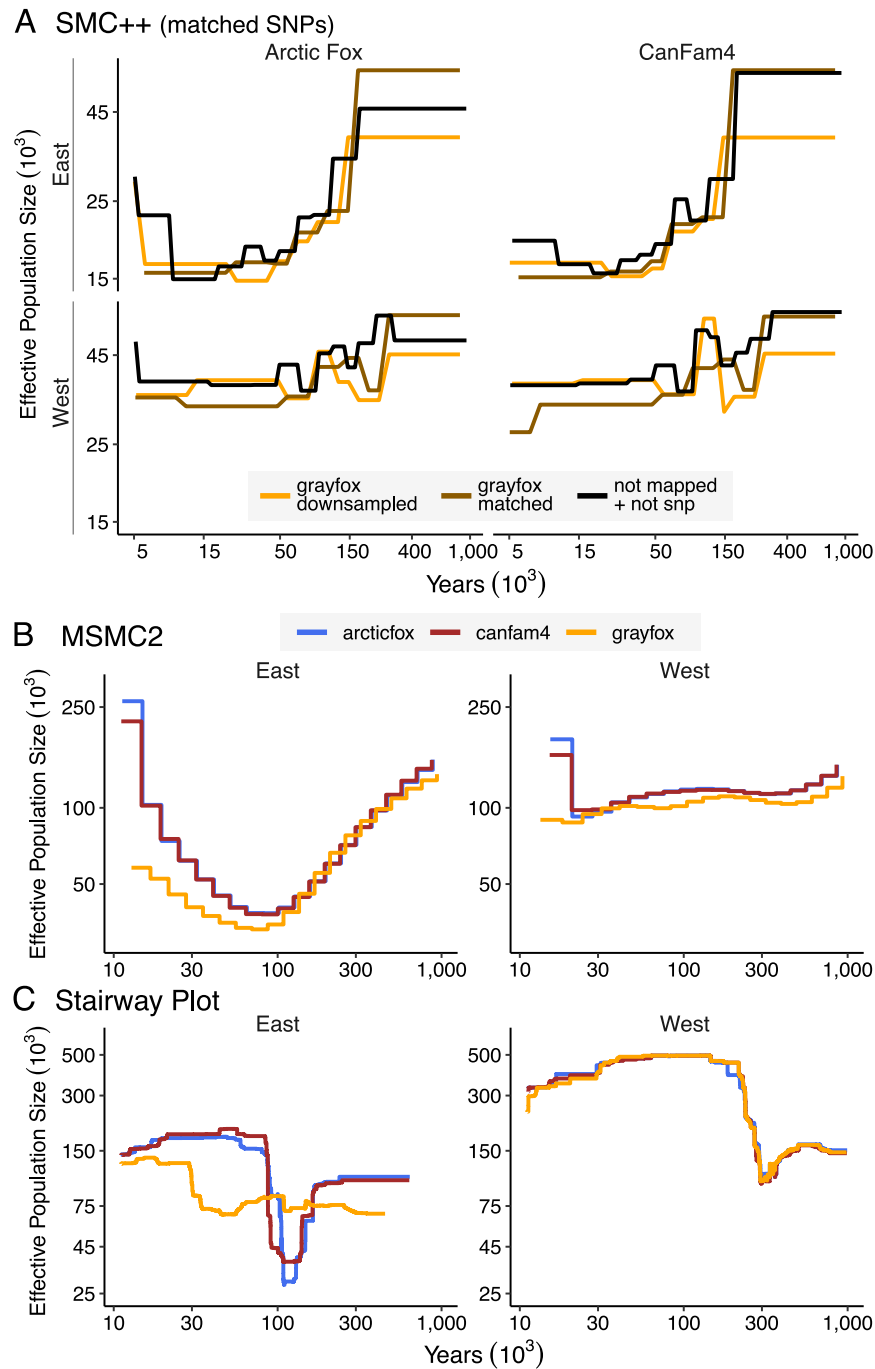


Figure S2. Demographic trajectories using different datasets and methods, related to Figure 2

(A) Demographic trajectories using matched SNP sets across reference genomes. Inferred effective population sizes (y axis) over time in years from present (x axis) for eastern and western populations using Arctic fox and CanFam4 reference genomes. Black lines: trajectories based on heterospecific genomes using only lifted-over SNPs (invariant sites and unmapped sites masked). Brown lines: trajectories based on gray fox genome using only SNPs that successfully lifted over to the respective heterospecific genome. Gold lines: trajectories based on gray fox genome with randomly downsampled SNPs matching the number of variants that lifted over from each heterospecific genome.

(B and C) Demographic trajectories across reference genomes. 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 (B) MSMC2 and (C) stairway plot.

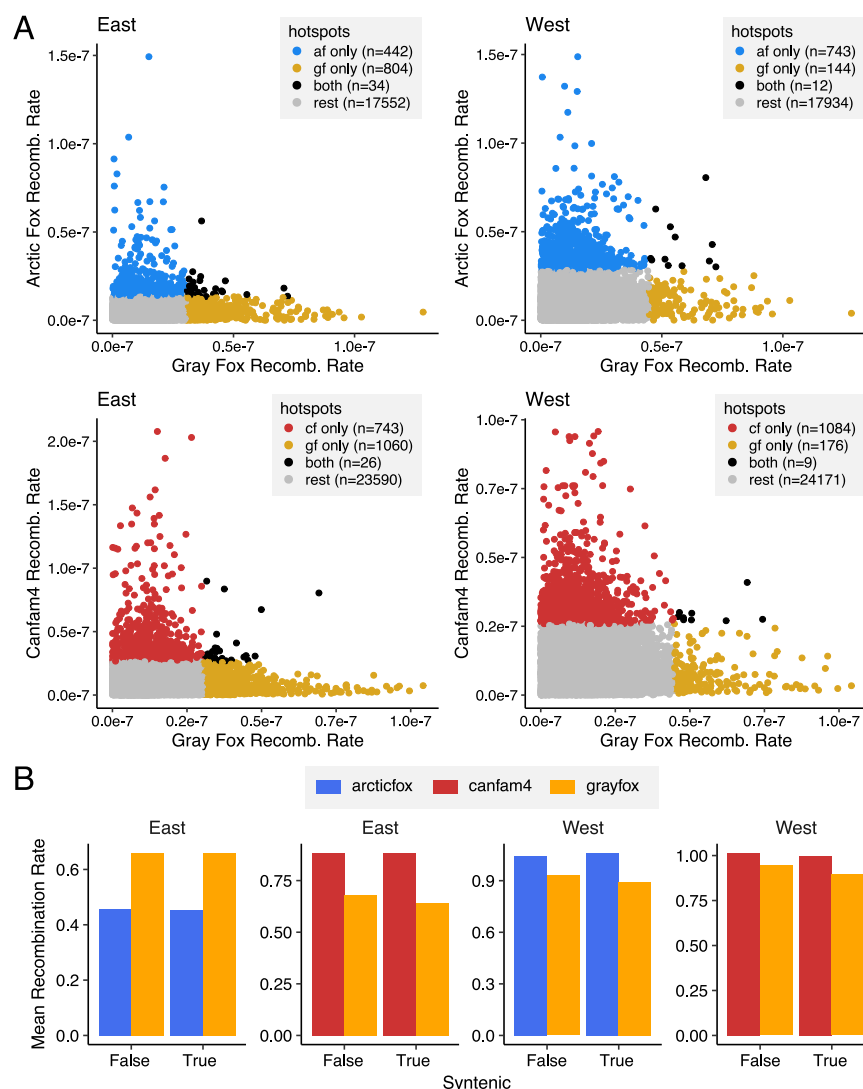


Figure S3. Reference genome choice affects recombination rate inference and hotspot detection, related to Figure 3

(A) Reference genome choice influences recombination hotspot detection. Scatterplots comparing recombination rates in 50-kb windows between gray fox and heterospecific references for eastern (left) and western (right) populations. Recombination hotspots unique to the arctic fox (blue, top), Canfam4 (red, bottom), and gray fox (gold) are highlighted. Hotspots shared between the two references are shown in black, and the remaining windows are in gray. The limited overlap in hotspots across references, with heterospecific references identifying substantially more unique hotspots in the western population but fewer in the eastern population compared with the conspecific gray fox reference, demonstrates that reference genome selection critically affects inferred recombination landscapes.

(B) Mean recombination rates for syntenic vs. non-syntenic chromosomes between reference genomes. Recombination rates vary more across references than between syntenic and non-syntenic regions.

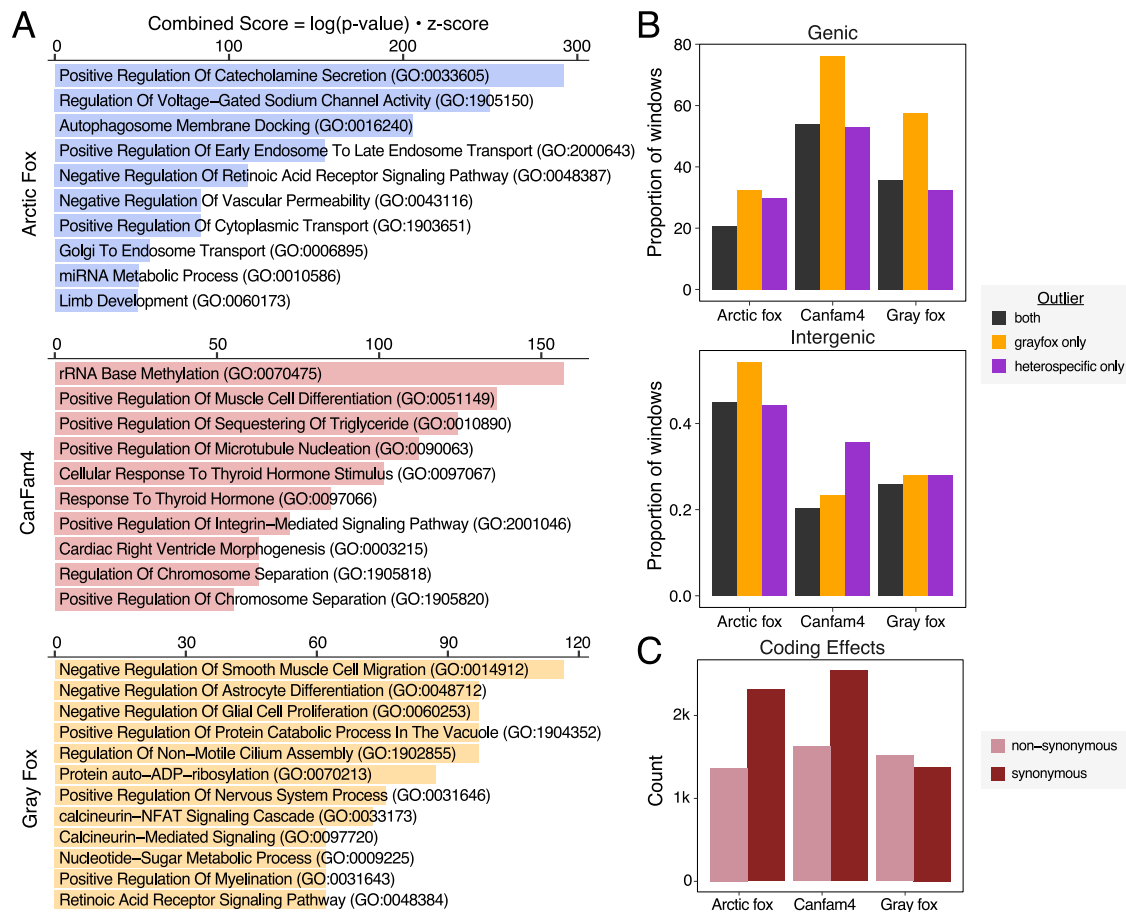


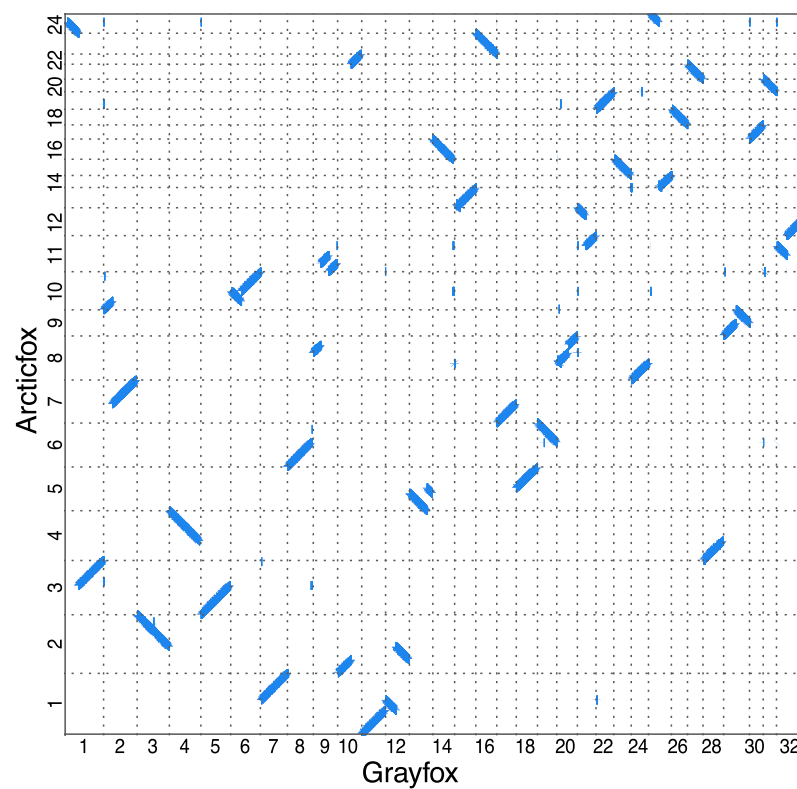
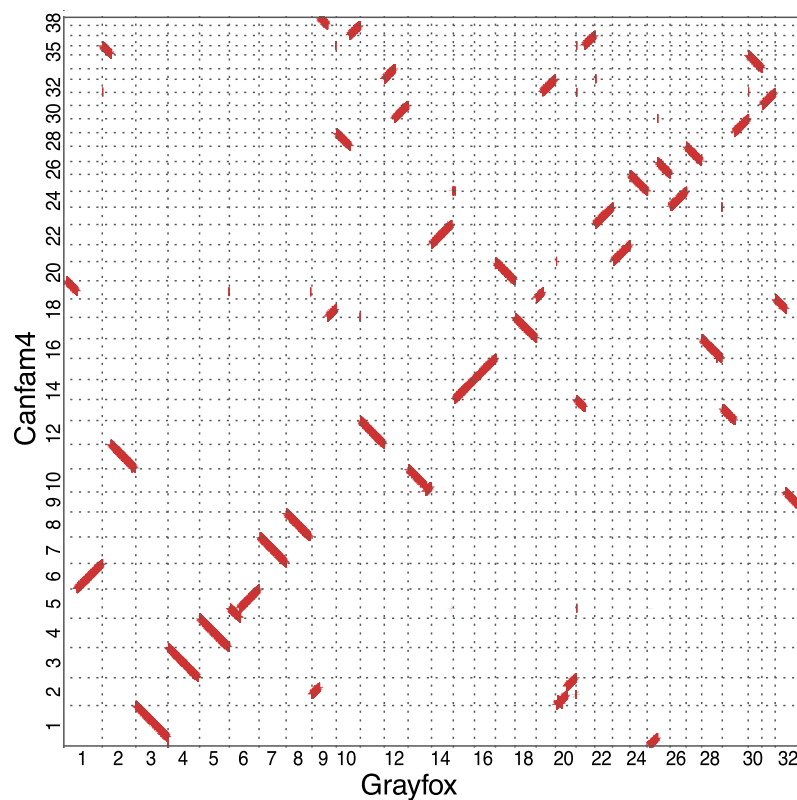
Figure S4. Reference genome choice affects functional annotation and interpretation of F_{ST} outliers, related to Figures 4 and 5

(A) 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.

(B and C) Functional context of F_{ST} outliers by reference genome.

(B) Genic (top) and intergenic (bottom) content of outlier windows based on annotations for the three references, with shared outliers (present in both the heterospecific and conspecific genomes) shown in black, outliers unique to gray fox shown in gold, and outliers unique to Arctic fox or Canfam4 shown in purple. Genic regions are consistently overrepresented in outliers across all references, with the highest enrichment in gray fox-specific outliers, while intergenic regions show variable representation depending on genome annotation.

(C) Number of non-synonymous (pink) and synonymous (red) mutations in outlier windows based on each reference. Heterospecific references show nearly twice as many synonymous as non-synonymous mutations, while the conspecific gray fox reference shows the opposite pattern, likely reflecting differences in gene models rather than variant detection, as total mutation counts remain similar across references.



(legend on next page)

Figure S5. Dot plot of pairwise whole-genome alignments, related to [Figure 1](#)

Plots are between the gray fox genome (x axis) and the heterospecific reference genomes (y axis, top) Canfam4 and the Arctic fox (y axis, bottom) showing synteny is not conserved, with evidence of chromosomal fusions and fissions.