

## Running title: Pleistocene origin of the red wolf

### Pleistocene origins, western ghost lineages, and the emerging phylogeographic history of the red wolf and coyote

Benjamin N. Sacks<sup>a,\*</sup>, Kieren J. Mitchell<sup>b</sup>, Cate B. Quinn<sup>a</sup>, Lauren M. Hennelly<sup>a</sup>, Mikkel-Holger S. Sinding<sup>c</sup>, Mark J. Statham<sup>a</sup>, Sophie Preckler-Quisquater<sup>a</sup>, Steven R. Fain<sup>d</sup>, Logan Kistler<sup>e</sup>, Stevi L. Vanderzwan<sup>a</sup>, Julie A. Meachen<sup>f</sup>, Elaine A. Ostrander<sup>g</sup>, Laurent A. F. Frantz<sup>h</sup>

<sup>a</sup>Mammalian Ecology and Conservation Unit/Veterinary Genetics Laboratory and Dept. Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, One Shields Avenue/Old Davis Rd., Davis, California 95616; <sup>b</sup>Australian Centre for Ancient DNA (ACAD) and ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), School of Biological Sciences, University of Adelaide, North Terrace Campus, South Australia, 5005 Australia; <sup>c</sup>Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen N, Denmark; <sup>d</sup>National Fish & Wildlife Forensic Laboratory, 1490 E Main St, Ashland, OR 97520; <sup>e</sup>Department of Anthropology, Smithsonian Institution, National Museum of Natural History, Washington, DC 20013; <sup>f</sup>Anatomy Department, Des Moines University, Des Moines, IA 50312; <sup>g</sup>National Human Genome Research Institute, National Institutes of Health, Bethesda MD, 20892; <sup>h</sup>Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University of Munich, Munich, Germany, School of Biological and Chemical Sciences, Queen Mary University of London, London, United Kingdom

BNS (0000-0003-0143-6589), KJM (0000-0002-3921-0262), CBQ (0000-0002-7685-362X), LMH (0000-0002-8805-2504), M-HSS (0000-0003-1371-219X), MJS (0000-0002-5416-8699), SP-Q (0000-0001-8885-5297), SRF (0000-0002-9748-7745), LK (0000-0002-5730-5986), SLV (0000-0002-7400-6961), JAM (0000-0002-2526-2045), EAO (0000-0001-6075-9738), LAFF (0000-0001-8030-3885)

#### Corresponding Author information

\*Benjamin N. Sacks, Mammalian Ecology and Conservation Unit, Veterinary Genetics Laboratory, Dept. Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, One Shields Avenue/Old Davis Rd., Davis, California 95616-8744 USA

Email: [bensacks@ucdavis.edu](mailto:bensacks@ucdavis.edu)

#### Abstract

The red wolf (*Canis rufus*) of the eastern US was driven to near-extinction by colonial-era persecution and habitat conversion, which facilitated coyote (*C. latrans*) range expansion and widespread hybridization with red wolves. The observation of some gray wolf (*C. lupus*) ancestry within red wolves sparked controversy over whether it was historically a subspecies of gray wolf with its predominant “coyote-like” ancestry obtained from post-colonial coyote hybridization (2-species hypothesis) versus a distinct species closely related to the coyote that hybridized with gray wolf (3-species hypothesis). We analyzed red wolf and other mitogenomes sourced from before the 20th century bottleneck and coyote

invasion, along with hundreds of modern amplicons, which led us to reject the 2-species model and to investigate a broader phylogeographic 3-species model suggested by the fossil record. Our findings broadly support this model, in which red wolves ranged the width of the American continent prior to arrival of the gray wolf to the mid-continent ~60 ka; red wolves subsequently disappeared from the mid-continent, relegated to California and the eastern forests, which ushered in emergence of the coyote in their place (~30 ka); by the early Holocene (12–10 ka), coyotes had expanded into California, where they admixed with and phenotypically replaced western red wolves in a process analogous to the 20<sup>th</sup> century coyote invasion of the eastern forests. Although nuclear data from historical and ancient specimens are needed to further investigate and refine this hypothesis, including the role of gray wolf admixture in the recent evolution of red wolves, findings indicate that the red wolf predated not only European colonization but humans in North America and suggest they predated coyotes as well. These findings highlight the urgency of expanding conservation efforts for the red wolf.

## Keywords

*Canis*, evolution, hybridization, mitogenome, phylogeography, red wolf

## 1 | Introduction

European settlement of North America was characterized by a dramatic increase in invasive species, novel diseases, accelerated extractive and agricultural land-use, and persecution of top carnivores. These practices transformed natural communities before they could be adequately understood and documented (Cronon, 1983; Nowak, 1979). One emblematic example is that of the wolf-like canids of the eastern forests: the red wolf (*Canis rufus*) of the USA and the eastern wolf (*C. lycaon*) of southeastern Canada (Nowak, 2002). Combined with destruction of their forested habitats, persecution of these indigenous canids reduced their numbers in the 1930s to the point where the coyote (*Canis latrans*), a smaller, opportunistic competitor, was able to expand its range into the eastern forests (McCarley, 1962; Hody and Kays, 2018). By 1975, coyotes had hybridized extensively with red and eastern wolves, helping to drive them to the point of near-extinction. This, in turn, prompted capture of the last known wild red wolves in an effort to preserve them through captive breeding and reintroduction programs (Kolenosky, 1971; Nowak, 1979; Phillips, Smith, Henry, & Lucash, 1995).

The genetic bottleneck caused by the collapse of the red wolf population, along with potentially extensive coyote introgression, severely compromised the reliability of modern red wolf genomes for understanding the evolutionary origins of the species (Heppenheimer, Brzeski, et al., 2018; McCarley, 1962; Murphy, Adams, Cox, & Waits, 2019; Phillips et al., 1995; vonHoldt et al., 2016). Modern red wolves, which descend from 12 of 14 captive founders, exhibit admixed genomes, composed primarily of coyote-like ancestry, but with significant gray wolf-like ancestry as well (Sinding et al., 2018; vonHoldt et al., 2016; vonHoldt et al., 2011), and only a single, private mitochondrial D-loop haplotype (Wilson et al., 2000). Given uncertainties surrounding the genetic heritage of red wolves prior to their extirpation from the wild, it is unclear whether their composite ancestry is the result of recent or ancient admixture, or both. This timeframe is important to conservation policy because the presumption that the red wolf originated in the 20<sup>th</sup> century as a hybrid facilitated by anthropogenic actions has fueled opposition to red wolf conservation (Gittleman and Pimm 1991; National Academies of Sciences and Medicine 2020).

Although a small population of eastern wolves persists in the wild, similar uncertainties surround their ancestry, which are exacerbated by extensive contemporary hybridization with the larger gray wolf (*Canis lupus*) (Heppenheimer, Harrigan, et al., 2018; Kolenosky, 1971; Rutledge, Devillard, Boone, Hohenlohe, & White, 2015; Wheeldon & White, 2009; Wilson, Grewal, Mallory, & White, 2009; Wilson, Rutledge, Wheeldon, Patterson, & White, 2012). The eastern wolf range was historically continuous with the red wolf range and is currently restricted to a region that was under glaciers before the early

Holocene. Thus, red and eastern wolves reflect geographically distant remnants of a once continuously distributed entity (e.g., Wilson et al., 2000). We hereafter use the term “red wolf” to refer collectively to both of these mid-sized *Canis* populations, using the separate terms when necessary for clarity; we use size-based terms, “mid-sized” and “small”, to refer to phenotypic red wolves and coyotes, respectively, without phylogenetic implications.

As a result of the dramatic anthropogenic impacts to the genetic composition of red wolves, two diametrically opposed hypotheses regarding their ancestry persist, fueling vigorous debate over conservation policy and management (Chambers, Fain, Fazio, & Amaral, 2012; National Academies of Sciences & Medicine, 2019, 2020; Waples, Kays, Fredrickson, Pacifici, & Mills, 2018). Most contentious has been whether the red wolf reflects a subspecies of gray wolf inhabiting the eastern forests that hybridized with coyotes during the late colonial era (Lawrence & Bossert, 1967; Lehman et al., 1991; vonHoldt et al., 2016; Wayne & Jenks, 1991) versus a single, ancient indigenous American wolf, distinct from the gray wolf and more closely related to the coyote (Kyle et al., 2006; Rutledge et al., 2015; Wilson et al., 2000). While hybridization must have occurred at least once at some point in the history of red wolves (Heppenheimer et al. 2018; Sinding et al. 2018; vonHoldt et al. 2016), the most immediate question relates to the timing of red wolf differentiation from coyotes and gray wolves (anthropocene versus ancient), rather than the role of hybridization *per se*. Appropriately valuing the taxonomic significance of these entities also requires a general understanding of how their phenotypes correspond to their phylogeographic history.

Fossil evidence has been subject to differing interpretations that essentially reflect those about distinctiveness of historical (and modern) red wolves (i.e., 2-species vs. 3-species models). As far back as the 1800s, the dominant view was that red wolves were distinct from both coyotes and gray wolves (3 species hypothesis). Correspondingly, Young & Goldman (1944) considered mid-sized *Canis* remains from the last Pleistocene glacial cycle (Wisconsinan), spanning the continent from eastern forests to California, to be red wolves, except for the small *Canis* (i.e., coyotes) that showed up in western locations at the terminal Pleistocene and early Holocene. In California, small *Canis* remains did not appear in the fossil record until 12–10 ka, although they occurred in potentially earlier Wisconsinan assemblages of the mid-continent (e.g., Arizona; Nowak, 1979; Meachen & Samuels, 2012). Young & Goldman’s interpretation that Pleistocene mid-sized *Canis* from California to Florida were red wolves remains the most parsimonious of the 3-species hypotheses.

Independently of the late 20<sup>th</sup> century controversy over distinctiveness of the red wolf, the 2-species view has dominated interpretation of western Pleistocene mid-sized *Canis* remains as those of a large coyote (e.g., Kurtén 1974). This view has fueled interpretation of its disappearance from California at the start of the Holocene as an evolutionary transition to the modern coyote brought on by rapid in-situ adaptation to faunal and climatic changes of the time (Meachen & Samuels, 2012). However, it is alternatively possible and more consistent with the 3-species model that the mid-sized *Canis* of the West was a red wolf (Young & Goldman, 1944) that was replaced through competition and admixture from an expanding population of coyotes analogous to the 20<sup>th</sup> century near-replacement of eastern red wolves by coyotes.

Ultimately, the fossil record alone is insufficient to assess whether the mid-sized *Canis* that predominated much of Wisconsinan mid-latitude America was a distinct entity (i.e., red wolf) versus a large version of the modern coyote or a small version of the modern gray wolf. What is clear, however, is that during historical times eastern red wolves and coyotes reflected two phenotypically distinct entities that were never considered conspecific. Thus, if red wolves were not a small or recently hybridized gray wolf (2-species hypothesis), they must have diverged from coyotes prior to European colonization, although not necessarily hundreds of thousands of years ago as is sometimes suggested

(Wilson et al., 2000; Waples et al., 2018). Further, if the coyote and red wolf represent a more recent, e.g., Wisconsinan-age, divergence, the question arises whether red wolf or coyote phenotypes represent the newer phenotype and, correspondingly, which entity more faithfully represents the original pre-gray wolf mid-latitude North American *Canis*.

In this study, we used complete mitochondrial genomes from historical red and eastern wolf specimens that predate the 20<sup>th</sup> century bottleneck and eastward expansion of coyotes, along with mitogenomes and amplicons from other historical and modern *Canis* specimens (Fig. 1; Tables S1–S4), to address these questions. To date, no phylogeny based on complete mitogenomes has been constructed for North American *Canis*, including coyotes. Short mitochondrial amplicons (200–500 bp) from historical red wolf specimens indicate a predominance of “coyote-like” matrilineal ancestry (Brzeski, DeBiasse, Rabon Jr, Chamberlain, & Taylor, 2016; Leonard & Wayne, 2008; Roy, Geffen, Smith, & Wayne, 1996; Rutledge, Bos, Pearce, & White, 2010; Wayne & Jenks, 1991; Wheeldon & White, 2009; Wilson, Grewal, McFadden, Chambers, & White, 2003) and incomplete lineage sorting between red wolves and coyotes, suggesting that divergence between red wolves and coyotes was unlikely to pre-date the last Pleistocene glaciation (i.e., Wisconsinan). However, existing mitochondrial amplicon data comprise mainly short fragments of the D loop, which exhibits considerable homoplasy confounding evolutionary inferences (Goddard, Statham, & Sacks, 2015; Ingman & Gyllensten, 2001). These amplicons are therefore inadequate on their own to reconstruct robust topologies or to calibrate nodes and estimate divergence times.

Thus, while nuclear DNA from historical samples is ultimately necessary to resolve certain questions involving hybridization, our objective here was to narrow the possibilities. In particular, we sought to assess whether red wolves of the eastern forests could have originated specifically as a colonial-era hybrid and, if not, to investigate whether modern California coyotes could reflect admixture from an extinct western red wolf and an expanding mid-continent coyote population that functionally replaced it 12–10 ka. Including gray wolf mitogenomes in our analysis, we further investigate concordance of phylogeographic patterns relative to the timing of gray wolf arrival to the mid-continent.

## **2 | Materials and Methods**

### **2.1 | Samples**

For complete mitogenomes, we used samples from 15 museum specimens, 1 archeological specimen from a pre-Columbian site, and 27 modern tissue samples from the archive of the Mammalian Ecology and Conservation Unit at UC Davis (Fig. 1A; Table S1, S2). Most historical samples dated between 1854 and 1930. Mitogenome analyses also utilized 6 published coyote mitogenomes as well as 8 more coyotes, along with 3 eastern wolves from Algonquin Provincial Park, and 3 modern red wolves, assembled from 14 short-read datasets on GenBank. To enable calibration of North American gray wolf clades, we also included 92 additional gray wolf and dog mitogenomes, African wild dog (*Lycaon pictus*), dhole (*Cuon alpinus*), and other *Canis* species (SI Appendix, Table S2). Lastly, we sequenced 425 coyotes at portions of the cytochrome b gene (369 bp) and D-loop region (278 bp), which we analyzed alongside previously published amplicons from historical red and eastern wolves (Fig. 1B; Tables S3).

### **2.2 | Laboratory procedures**

Extraction of DNA from ancient and museum samples followed strict ancient and historical DNA protocols (Ameen et al., 2019; Aubry, Statham, Sacks, Perrine, & Wisely, 2009; Brown, Darwent, & Sacks, 2013; Reponen, Brown, Barnett, & Sacks, 2014), and was performed at an ancient DNA (aDNA) facility at UC Davis. Extracted DNA was shipped to myReads NGS services at Arbor Biosciences (Ann Arbor, MI, USA) for library preparation using ancient DNA protocols, pooling, and Illumina PE150 sequencing on a NovaSeq S4 lane at Novogene (Sacramento, CA).

We extracted DNA from modern samples manually using manufacturer's protocols (DNeasy Blood & Tissue Kit, Qiagen, Luxembourg). We used a capture enrichment protocol for library construction and sequenced mitogenome-enriched libraries on an Illumina PE250 MiSeq lane at the UC Davis Genome Center.

§ We assembled mitogenomes de-*Novo* using NovoPlasty (version 2.7.2, Dierckxsens, Mardulyn, & Smits, 2017) after adapter-trimming, quality filtering (discarding reads with >50% of bases with quality scores <2), and duplicate removal. For mitogenomes that did not result in circularized assemblies, we manually aligned reads aided by a coyote reference mitogenome in Sequencher (v 5.4.6; Gene Codes, Ann Arbor MI), calling positions with  $\geq 3\times$  coverage, substituting with Ns otherwise. We PCR-amplified and then Sanger-sequenced *cytb* and D loop fragments, resulting in trimmed sequences of 369 bp and 278 bp, respectively (Supplemental Methods).

### 2.3 | Phylogenetic and network analyses

Phylogenetic trees were constructed from the first 15,457 bases of the mitogenome (i.e., all but the D loop region) in Beast v 1.8.3 (Drummond & Rambaut, 2007) on the Cipres Science Gateway v 3.3 (Miller, Pfeiffer, & Schwartz, 2010). We constructed one tree based solely on 3rd codon positions to calibrate nodes (Subramanian & Lambert, 2011) and a second using 4 partitions corresponding to codon positions and noncoding bases to obtain the most highly resolved topology. Both trees were rooted to the African wild dog with a normal root height prior of 3.9 Ma (SD = 0.3 Ma) based on fossil calibrated nuclear DNA dating (Chavez et al., 2019). We used a birth-death model and relaxed lognormal clock, selected over a strict clock on the basis of its coefficient of variation (0.337, 95% HPD: 0.194–0.490, in the 3rd-codon position tree). Although we only specified a prior on the root height, we verified the consistency of internal nodes based on comparison to nuclear and mitochondrial calibrations (Chavez et al. 2019; Koepfli et al. 2015; Perini et al. 2010) and the closure of the Bering Strait  $\sim 12$  ka. We used 50 million MCMC cycles sampled every 5,000 cycles, which was sufficient for all parameters to achieve ESS > 200. Additional details can be found in the Supplemental Methods.

To calibrate the mutation rates of amplicons to those estimated for mitogenomes, we built a tree based solely on 647 bp of the cytochrome *b* (369 bp) and D-loop (278 bp) regions. This resulted in estimates of  $1.85 \times 10^{-8}$  substitutions per site per year (95% HPD: 0.71–3.62  $\times 10^{-8}$ ) for *cytb*,  $13.22 \times 10^{-8}$  substitutions per site per year (5.74–24.65  $\times 10^{-8}$ ) for D loop, and  $6.74 \times 10^{-8}$  substitutions per site per year (2.87–12.66  $\times 10^{-8}$ ) for the 647 bp combined sequence. Additional technical details are provided in Supplemental Methods. We built median-joining networks using Networks v 5.0 (Bandelt, Forster, Sykes, & Richards, 1995). The ages of network clades were determined from rho estimates, which are equivalent to the average number of derived mutations since divergence from the most recent common ancestor (Macaulay, Soares, & Richards, 2019).

### 2.4 | IMA analyses

We fitted Isolation-only and Isolation-with-Migration models using IMA2 to estimate splitting times (Hey & Nielsen, 2007). This approach was designed specifically for use with recently diverged populations, such as incipient species, for cases characterized by incomplete lineage sorting and potentially continued gene flow. We used likelihood ratio tests to assess whether models incorporating gene flow were warranted over those that did not. All results were replicated with a distinct random number seed, yielding consistent posteriors for splitting time. Posteriors for splitting time and migration achieved unimodal distributions approaching zero at both ends for all temporal estimates. Scaled parameters were converted to absolute values based on mutation rates described above. Additional technical details are provided in Supplemental Methods.

### 2.5 | Historical demography

To assess the historical demography reflected in mitogenomes, Bayesian skyline plots (Drummond, Rambaut, Shapiro, & Pybus, 2005) were constructed in BEAST (v1.8.3) using parameters and priors described above. Plots were constructed separately for subsets of the data to allow contrasting demographic histories. We generated 10 to 50 million MCMC cycles, as necessary to achieve stationarity and all parameters at ESS > 200. We also computed coalescent summary statistics indicative of population expansion using DNAsp 6.12.03 (Rozas et al., 2017).

### 3 | Results

#### 3.1 | Mitogenome phylogeny

We successfully sequenced full ( $n = 10$ ) or partial ( $n = 1$ ) mitogenomes from 11 of 16 historical *Canis* samples, comprising six red wolves (1904, 1905, 1916, 1917, 1924, 1941), three mid-continent coyotes (<1400, 1894, 1895), and two putative gray wolves (1893, <1913) (Fig. 1A; Table S1). The depth of coverage ranged from 20x to 2,054x after removal of duplicates (Tables S1, S2). We also produced mitogenomes from 145 modern canids (Table S2).

We constructed a phylogenetic tree from the 155 mitogenomes in BEAST (Fig. 2A–C, Fig. S1). All red wolf, eastern wolf, and coyote mitogenomes, as well as one mitogenome from a historical putative gray wolf from Nebraska, which is hereafter considered a red (or eastern) wolf matriline, formed a well-supported (BPP = 1.0) monophyletic clade that is endemic to North America (“American *Canis* clade”; Fig. 2C). This American *Canis* clade was reciprocally monophyletic with respect to gray wolves (Fig. 2A).

As noted in previous studies (Koblmüller et al., 2016; Loog et al., 2020; Weckworth, Talbot, & Cook, 2010), North American gray wolves comprised three monophyletic subclades nested within an otherwise Eurasian gray wolf clade (Fig. 2B). These three North American subclades shared MRCA with Eurasian gray wolves dating to 66 ka (95% HPD: 40–99 ka), 30 ka (17–46 ka), and 8 ka (4–15 ka). The two older lineages included Mexican gray wolves and an 1893 gray wolf specimen from Colorado, respectively. These TMRCA reflect the earliest times that ancestors of extant North American gray wolves could have arrived in North America and are consistent with the lack of fossil evidence for the presence of gray wolves in mid-latitude North America (i.e., south of the ice sheets) until the mid-Wisconsinan glaciation (Meachen, Brannick, & Fry, 2016; Nowak, 2002).

As expected, coyote and red wolf matriline were not reciprocally monophyletic within the American *Canis* clade (Fig. 2C). The American *Canis* clade was divided into two deeply divergent (TMRCA = 209 ka, 123–330 ka) reciprocally monophyletic clades, hereafter referred to as clades A and B (Fig. 2C). Clade A contained only California coyotes. Clade B was further divided into reciprocally monophyletic clades b1 and b2. Except for a coyote from the Texas hybrid zone, mid-continent coyote haplotypes fell into the youngest clade, b2, with most (75%) of them falling within two subclades (b2i, b2ii) that also contained the two pre-1900 mid-continent coyotes. These coyote subclades post-dated the earliest divergence of a North American gray wolf lineage from its Eurasian counterpart (node f, in Figs. 2B, D), consistent with a possible link between coyote expansion and gray wolf arrival to the mid-continent.

#### 3.2 | Could the red wolf be a gray wolf admixed with coyotes during the colonial era?

Our results revealed no evidence for gray wolf mitochondrial haplotypes within red wolves. Therefore, these red wolf matriline either reflect ancient red wolf ancestry or they descend recently from female coyotes that hybridized with male gray wolves during the post-colonial era. In the latter case (gray wolf co-option of coyote matriline), however, the mitochondrial diversity observed in red wolves would be expected to represent a small subset of modern mid-continent coyote diversity (Waples et al., 2018). Contrary to this prediction, the red and eastern wolf mitogenomes exhibited significantly greater nucleotide diversity (0.0041, SD = 0.00039,  $n = 12$ ) than that of coyotes of the mid-continent (0.0025, SD

= 0.0004,  $n = 16$ ) (Fig. 2E; Table S5A). This pattern was recapitulated by analyses of over 400 shorter mitochondrial sequences (Fig. 1B), which additionally included seven published historical red wolf *cytb* sequences (Roy et al., 1996) (Tables S3, S4; Fig. S2). For example, *cytb* nucleotide and haplotype diversity ( $\pm$ SD) were higher in the 15 red wolves ( $0.0057 \pm 0.0010$ ;  $0.81 \pm 0.078$ , respectively) than in the 189 mid-continent coyotes ( $0.0026 \pm 0.0003$ ;  $0.49 \pm 0.039$ ). We found only seven *cytb* haplotypes in mid-continent coyotes versus five haplotypes in the much smaller red wolf sample ( $n = 15$ ; Fig. S2); three of the five (60%) red wolf haplotypes were not found in mid-continent coyotes despite the coyotes' much larger sample size. The only two haplotypes shared between red wolves and coyotes were also the most basal, consistent with derived haplotypes reflecting post-divergence mutations (Fig. S2A).

Next, we used the mitochondrial genomes to estimate the time of divergence between red wolf and coyote matriline. Because the phenotypically distinct red wolves and coyotes did not correspond to reciprocally monophyletic clades, we used IMA2, which was designed specifically to take advantage of non-random coalescence to estimate divergence time between recently diverged groups with incomplete lineage sorting and that potentially exchange genes after divergence as well (Hey and Nielsen 2007). Using all coyotes, this analysis yielded an estimate of divergence between red wolves and coyotes dating prior to the last glacial maximum (LGM) at 32 ka (22–44 ka), with a small ( $m = 0.04$ , 95% HPD: 0.002–0.12) but significant ( $P < 0.05$ ) signal of post-divergence gene flow (Fig. 3A,B). The IMA2 result was consistent when California coyotes were excluded from the analysis (as mid-continent coyotes were the adjacent ones that could have hybridized with eastern red wolves) and was only incrementally affected by conservatively assuming absence of gene flow (Fig. 3B). To ensure that mid-continent coyotes were comprehensively represented, we conducted another IMA2 analysis using the 189 mid-continent coyote amplicons vs. red wolves. We calibrated the concatenated *cytb* and D loop haplotypes to our mitogenome tree using the same individuals for which we had complete mitogenomes; this analysis provided a splitting time estimate of 25 ka (95% HPD: 8–43 ka), which, although imprecise, was qualitatively consistent with that from the mitogenomes, and also revealed a significant signal of post-divergence gene flow ( $m = 0.76$ ; 95% HPD: 0.13–1.91,  $P < 0.0001$ ). We note that underestimates of gene flow correspond to underestimates of divergence in isolation-with-migration models (Hey and Nielsen 2007). Our near-zero estimates of gene flow in the mitogenome analysis were likely an underestimate given that gene flow is known to have occurred historically between coyotes and red wolves in Texas; e.g., Heppenheimer et al., 2020; Nowak, 2002; Fig. 1). Thus, our splitting time estimates also could be underestimated. For example, a Bayesian skyline plot constructed from mitogenomes suggested that trajectories of red wolves and mid-continent coyotes could have diverged as early as 50 ka or more, but still squarely within the Wisconsinan glaciation (Fig. 3C).

### 3.3 | Amplicon analysis of ghost lineages and recent admixture in California coyotes

The larger amplicon dataset contextualized with the mitogenome tree suggests that California coyotes maintain significant maternal ancestry from the mid-sized Pleistocene canids that disappeared from California 12–10 ka, but also share more recent ancestry with mid-continent coyotes (Fig. 4). In particular, clade A was nearly endemic to California coyotes, and clade b1, common to both California coyotes and red wolves, also was absent from mid-continent coyotes outside the Texas hybrid zone; only clade b2 was found frequently in all three groups and it composed 99% of all matriline in mid-continent coyotes. A haplotype network further revealed differences in the timing of shared ancestry among these three clades (A, b1, b2; Figs. 5A, S2). Only within clade b2 was haplotype sharing common and therefore consistent with relatively recent gene flow. In contrast, the few A-clade haplotypes observed in mid-continent coyotes were deeply divergent from all those in California, as were b1-clade haplotypes carried by California coyotes versus red (and eastern) wolves. Thus, the A and b1 ancestry in California coyotes, collectively 58% of their maternal ancestry, predates the phenotypic shift observed

in the California fossil record of 12–10 ka and therefore reflect ghost lineages from the larger Pleistocene predecessor. The sharing of clade b1 by California coyotes (i.e., presumed ghost lineages) and historical eastern red wolves further suggests that the mid-sized Pleistocene canids of both ends of the continent shared ancestry during the early to mid-Wisconsinan glaciation (rho estimate = 57 ka, SD = 26 ka; Figs. 5A, S2). Conversely, the many shared clade-b2 haplotypes between mid-continent coyotes and California coyotes indicate recent to contemporary gene flow.

To further investigate this admixture hypothesis, we quantified divergence time and directional gene flow in IMA2 using the 416 California and mid-continent coyote amplicons. This analysis indicated a Wisconsinan-age divergence (61 ka, 95% HPD: 12–99 ka) followed by asymmetrical gene flow predominantly from east to west as predicted if mid-continent coyotes contributed to the late Pleistocene transformation of the California mid-sized *Canis* through admixture (Fig. 5C). Consistent with asymmetrical gene flow from the mid-continent into California, coalescent statistics exhibited signatures of population expansion for mid-continent coyotes, including significantly negative values for Tajima's D (-1.75,  $P < 0.05$ ,  $n = 16$ ),  $D^*$  (-1.85,  $P < 0.05$ ), and  $F^*$  (-2.05,  $P < 0.05$ ) in the mitogenome dataset (Table S5A). Results were similar for amplicon datasets (Table S5B–D). By comparison, none of these statistics differed significantly from zero for the California coyote population (Tables S5A–D). Lastly, Bayesian skyline plots constructed from the high-resolution mitogenomes indicated a sudden rise in clade B matriline beginning at the end of the Pleistocene and start of the Holocene (Fig. 5B), approximately coincident with the phenotypic transformation of California *Canis*. Given that this early-Holocene pulse was only evident in Bayesian skyline plots with California coyotes included (i.e., not iFig. 3C) and regardless of whether A-clade haplotypes were included (Fig. 5B), it must partly reflect an increase in B-clade haplotypes in California.

## 4 | Discussion

### 4.1 | Could the red wolf be a colonial-era hybrid?

If the red wolf were a colonial-era hybrid, we should have observed in our sample of pre-bottlenecked red wolves either gray wolf matriline or coyote-like matriline reflecting a small subset of those found in coyotes (e.g., Waples et al., 2018). To the contrary, not only did we find no gray wolf matriline but the diversity of coyote-like matriline was actually higher in red (and eastern) wolves than in coyotes themselves, particularly from the mid-continent, both for our limited set of complete mitogenomes and in the much larger coyote amplicon data set. This diversity in pre-bottlenecked red wolves contrasts with the four modern post-bottlenecked red wolves, all of which shared essentially identical mitogenomes.

The only way to reconcile these findings with the hypothesis that red wolves represent colonial-era crossing of gray wolves with coyotes would be if a distinct group of coyotes from those sampled in this study existed and its matriline were completely absorbed into the red wolf. However, no phenotypic coyotes occurred in the eastern forests before mid-continent coyotes expanded into them, primarily in the early 20th century (Nowak 2002). Nor did coyotes in the mid-continent suffer a population decline that could explain such a massive loss of mitochondrial diversity over the past century. On the contrary, their populations expanded substantially since colonial times (Hody and Kays, 2018; McCarley, 1962). Our presumption that modern coyote diversity differs little from pre-colonial times was also supported by our two ancient mid-continent coyote mitogenomes, which clustered in the same two coyote clades (b2i, b2ii) that predominated in the modern mid-continent coyotes in this study (mitogenomes and amplicons). Finally, the IMA2 analyses, which indicated divergence between coyote and red wolf matriline prior to the LGM, provide quantitative evidence not only of a pre-colonial origin of the red wolf, but origin of the red wolf prior to any human presence in the Americas. Thus, our study provides



strong evidence that the red wolf represents an indigenous North American canid that evolved independently of anthropogenic influence.

#### **4.2 | Was the Pleistocene California mid-sized *Canis* a red wolf?**

Our results also lend support to Goldman's interpretation that mid-sized Pleistocene canid fossils from California were red wolves rather than coyotes (Young and Goldman, 1944). First was our finding that over half the matrilineal lineages carried by California coyotes (i.e., in clades A and b1) diverged from mid-continent coyote matrilineal lineages many thousands of years before small *Canis* existed in California; this observation (based on the mitogenome tree, amplicon networks, and IMA2 analyses) suggests that the A and b1 haplotypes in California coyotes descended from the Pleistocene mid-sized *Canis* of California independently of the ancestry of mid-continent coyotes, which primarily reflects the more recently derived b2 clade. Second, our finding that red and eastern wolves, but not mid-continent coyotes, also carried haplotypes from one of the California clades (b1) implies past connectivity between the Pleistocene California *Canis* and red wolves. This connectivity could have occurred no earlier than the TMRCA of the b1 clade, which the mitogenome tree and amplicon data (rho estimates) place squarely within the Wisconsinan glaciation, ~50–60 ka. We found no A-clade haplotypes (the most common in California coyotes) in red wolves, however, possibly due to our small historical red wolf sample size. Given the phenotypic similarity between Pleistocene California *Canis* and red wolves, however, such recent (~50–60 ka) connectivity as demonstrated from the b1 clade is most parsimoniously interpreted as conspecificity. Thus, in agreement with Young and Goldman (1944), we suggest that Pleistocene California *Canis* was most likely a red wolf population.

Regardless of one's taxonomic classification of the mid-sized Pleistocene *Canis* of the West, our findings provide strong evidence of recent admixture from mid-continent coyotes coinciding with its phenotypic transformation. Indeed, in stark contrast with A and b1 haplotypes, many of the b2 haplotypes in California were shared with mid-continent coyotes, indicating recent gene flow. The IMA2 analysis supported a model of Wisconsinan-age divergence between ghost matrilineal lineages of Pleistocene California *Canis* versus mid-continent coyotes followed by more recent asymmetrical gene flow dominated by that from the mid-continent into California. Along with coalescent statistics, which indicated a population expansion by mid-continent coyotes in particular, a Bayesian skyline plot revealed a pulse driven by an increase in b2 haplotypes in California at approximately the same time as the fossil record revealed the rapid reduction in body size of *Canis* in California (~12–10 ka). Thus, although we cannot rule out that admixture from mid-continent coyotes and the phenotypic transformation of California *Canis* was coincidental, the pattern bears an uncanny resemblance to the 20<sup>th</sup> century expansion of coyotes into the eastern forests, followed by hybridization and near-extirpation of red wolves (McCarley, 1962). Modern southeastern coyotes occupying the former range of the red wolf similarly carry ghost alleles from extirpated red wolves (Heppenheimer et al., 2018, 2020; Murphy et al., 2019; Sinding et al. 2018).

#### **4.3 | Could the coyote be a recently evolved offshoot of the red wolf?**

In accordance with the fossil record, we also found no evidence to support that the red wolf was a late-Pleistocene offshoot of a more ancient coyote ancestor, or even a symmetrical sister taxon to the coyote. In contrast, our data - interpreted alongside existing palaeontological data - seem more in line with a model whereby the coyote phenotype, rather than the red wolf phenotype, was the more recent evolutionary product, and the modern red wolf therefore more likely represents phenotypic continuity with their common ancestor. Although the mitochondrial haplotypes of mid-continent coyotes (99% belonging to the b2 clade) reflect high diversification associated with recent population expansion, the entire b2 clade (i.e., MRCA) traces back relatively recently to the mid-Wisconsinan glaciation (~50–60 ka, based on the mitogenome tree), corresponding approximately to its estimated divergence with red wolves based on Bayesian skyline plots. The TMRCA of the two primary coyote subclades (b2i, b2ii) are

even more recent, as were the IMA2 point estimates for the divergence between mid-continent coyotes and red wolves (~30 ka). The most persuasive evidence, however, comes from the fossil record, which indicates that mid-sized *Canis* dominated western and eastern ends of the continent and, although based on fragmentary evidence, the mid-continent as well, prior to the latter half of the Wisconsinan glaciation, when small *Canis* began to appear (Meachen & Samuels, 2012; Kurtén, 1974; Nowak, 1979). Thus, notwithstanding uncertainties surrounding the origins of the coyote, the red wolf could represent a much more ancient entity than suggested by its matrilineal divergence from the modern coyote.

Owing to the sparsity of the fossil record of the mid-continent, even sparser in Mexico and Central America, the circumstances and exact timing of the emergence of the coyote phenotype in mid-latitude North America remains unclear. Did the coyote evolve only 60–30 ka in-situ as an offshoot of mid-sized wolf-like canids? Alternatively, might the smaller *Canis* have existed further south (i.e., southern Mexico or central America) before its phenotype emerged in the fossil record of mid-latitude North America (e.g., *Canis hondurensis*, Goldman, 1936)? If so, did this putative small southern canid reflect one end of a continuous cline in body size, the mid-sized *Canis* of the mid-continent representing the other end? Or was it a phylogenomically distinct taxon that expanded north and hybridized with mid-sized *Canis*, co-opting a small subset of its matrilineal diversity? All three of these scenarios could be consistent with our mitochondrial data. Despite the historical occurrence of coyotes as far south as Costa Rica (and currently, Panama), we are aware of no DNA data from the southern portion of its range (Hidalgo-Mihart, Cantú-Salazar, González-Romero, & López-González, 2004; Hody & Kays, 2018; Lucas, Alvarado, & Vega, 1997). In the future, both mtDNA and nuclear genetic analyses of *Canis* from their southern range would help resolve these questions.

Based on the totality of our findings and the fossil evidence, we hypothesize that the arrival of the gray wolf to North America in the early to mid-Wisconsinan caused extirpation of red wolves from the mid-continent, which ushered in the expansion of coyotes (e.g., from the south or via in-situ evolution). In contrast to red wolves, coyotes and gray wolves were historically and are today sympatric throughout most of their North American ranges, where they occupy distinct ecological niches, and do not interbreed (REFs). Whatever the environmental factors that favored the coyote over red wolf phenotype in California at the end of the Pleistocene (Meachen and Samuels, 2012), the genetic underpinnings of this coyote phenotype likely originated from these newer expanding coyotes of the mid-continent. In the future, analyses of nuclear data from California and mid-continent coyotes, along with historical red and eastern wolves, could clarify the genomic landscape of this evolutionary transformation, as well as the potential role of gray wolf admixture in more recent red wolf evolution.

#### **4.4 | Conclusions and conservation implications**

Our study, made possible by the diligent preservation and curation of historical museum specimens, clarified several important aspects of the maternal phylogeography of North American *Canis*. Most significant were our findings indicating that the red wolf was not only a pre-colonial and, indeed, pre-human entity, but that its legacy likely predated that of the smaller-bodied coyote across mid-latitude North America. Our findings also leave open several questions in need of exploration with an expanded sample of historical and ancient specimens and with nuclear DNA. Chief among these questions are the role (e.g., adaptive vs incidental) and timing of hybridization with gray wolves and the evolutionary origins of the coyote.

As a consequence of persecution and hybridization, red wolves from the southeastern forests were driven to near-extinction before the last deemed “pure” red wolves were brought into captivity for rearing and eventual reintroduction to coastal North Carolina (Phillips et al., 1995). By the early 21<sup>st</sup> century, coyotes colonizing from the north and the south had met in the middle and occupied the entirety of the east coast (Hinton et al., 2019). Today, the descendants of red wolves, in their “purest”

form, are confined in the north to Algonquin Provincial Park in Ontario (eastern wolves, *C. lycaon*), where they are surrounded by admixed gray wolves (“Great Lakes wolves”) and coyotes, and in the south by the precarious reintroduced population of red wolves in coastal North Carolina and associated captive propagation facilities. Conservation policy and management of red and eastern wolves has been hampered by confusion and disagreement among scientists over their systematics and origins, owing in large part to the massive anthropogenic hybridization event of the 20<sup>th</sup> century, which has up to now obscured their evolutionary heritage. Our findings here highlight the invaluable diversity retained through existing recovery programs and underscore the urgency of policy changes and actions to safeguard these last remaining individuals and ensure persistence of the species.

Chavez DE, Gronau I, Hains T, Kliver S, Koepfli K-P, Wayne RK (2019) Comparative genomics provides new insights into the remarkable adaptations of the African wild dog (*Lycaon pictus*) Scientific reports 9:1-14

Gittleman JL, Pimm SL (1991) Crying wolf in North America Nature 351:524-525

Heppenheimer E et al. (2018) Population genomic analysis of North American eastern wolves (*Canis lycaon*) supports their conservation priority status Genes 9:606

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics Proceedings of the National Academy of Sciences 104:2785-2790

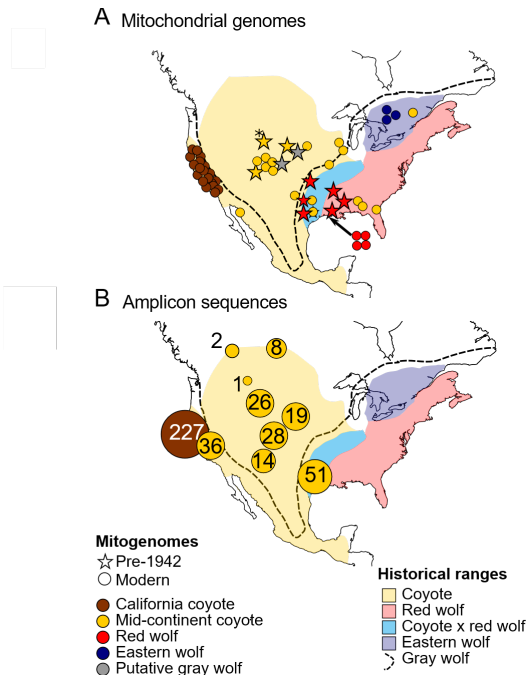
Koepfli K-P et al. (2015) Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species Current Biology 25:2158-2165

National Academies of Sciences E, Medicine (2020) A Research Strategy to Examine the Taxonomy of the Red Wolf. The National Academies Press, Washington, DC. doi:doi:10.17226/25891

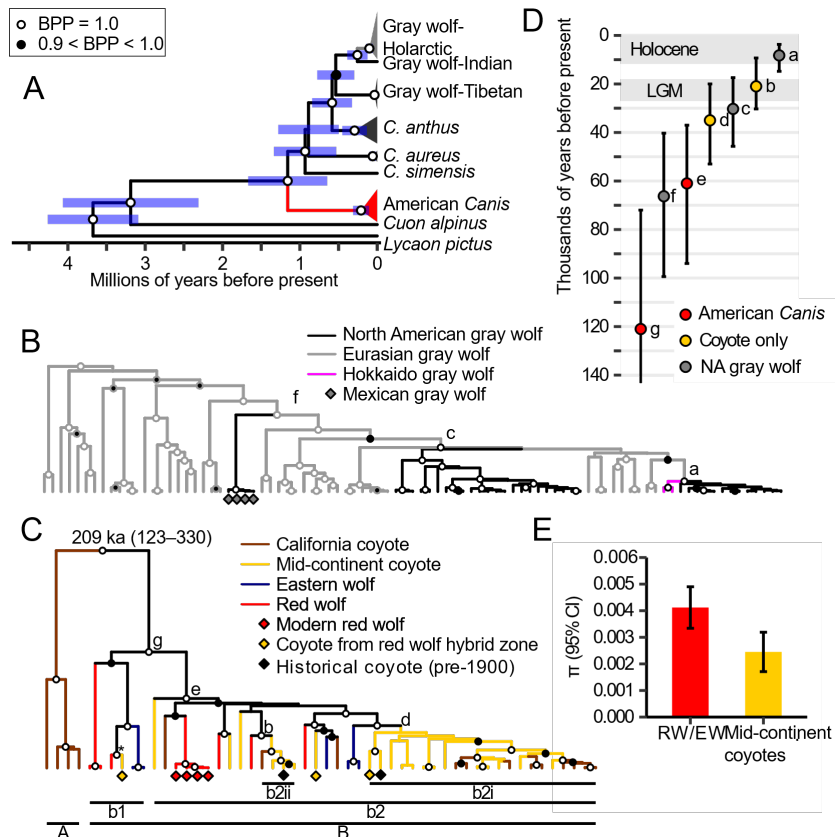
Perini FA, Russo C, Schrago CG (2010) The evolution of South American endemic canids: a history of rapid diversification and morphological parallelism Journal of evolutionary biology 23:311-322

Sinding M-HS et al. (2018) Population genomics of grey wolves and wolf-like canids in North America PLoS genetics 14:e1007745

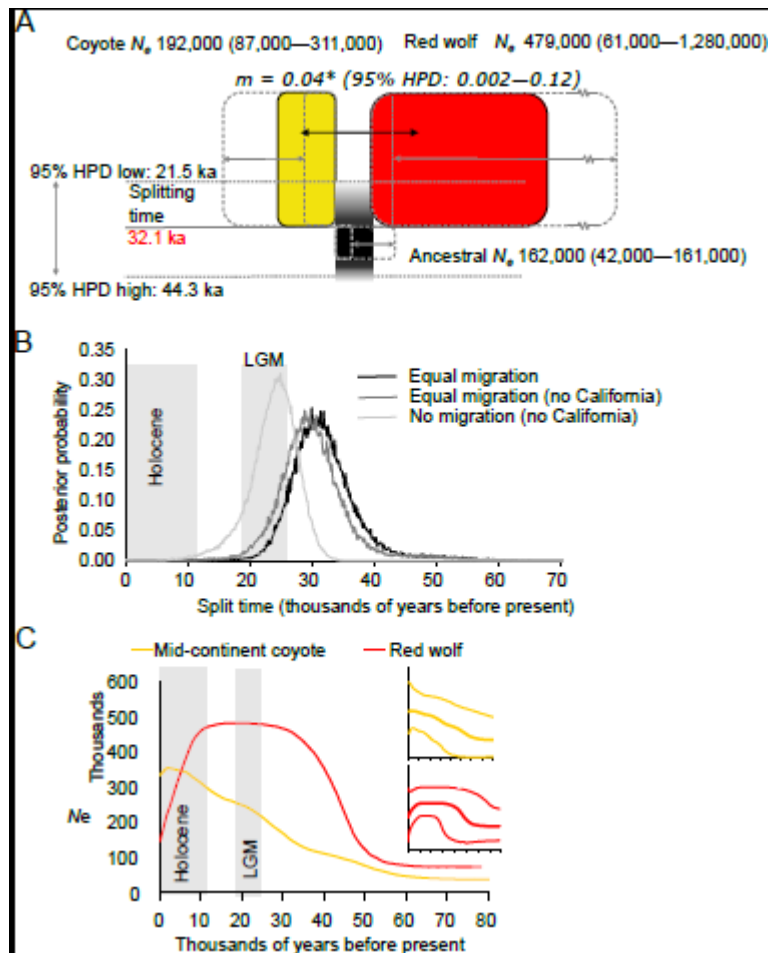
vonHoldt B et al. (2016) Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf Science Advances 2:e1501714



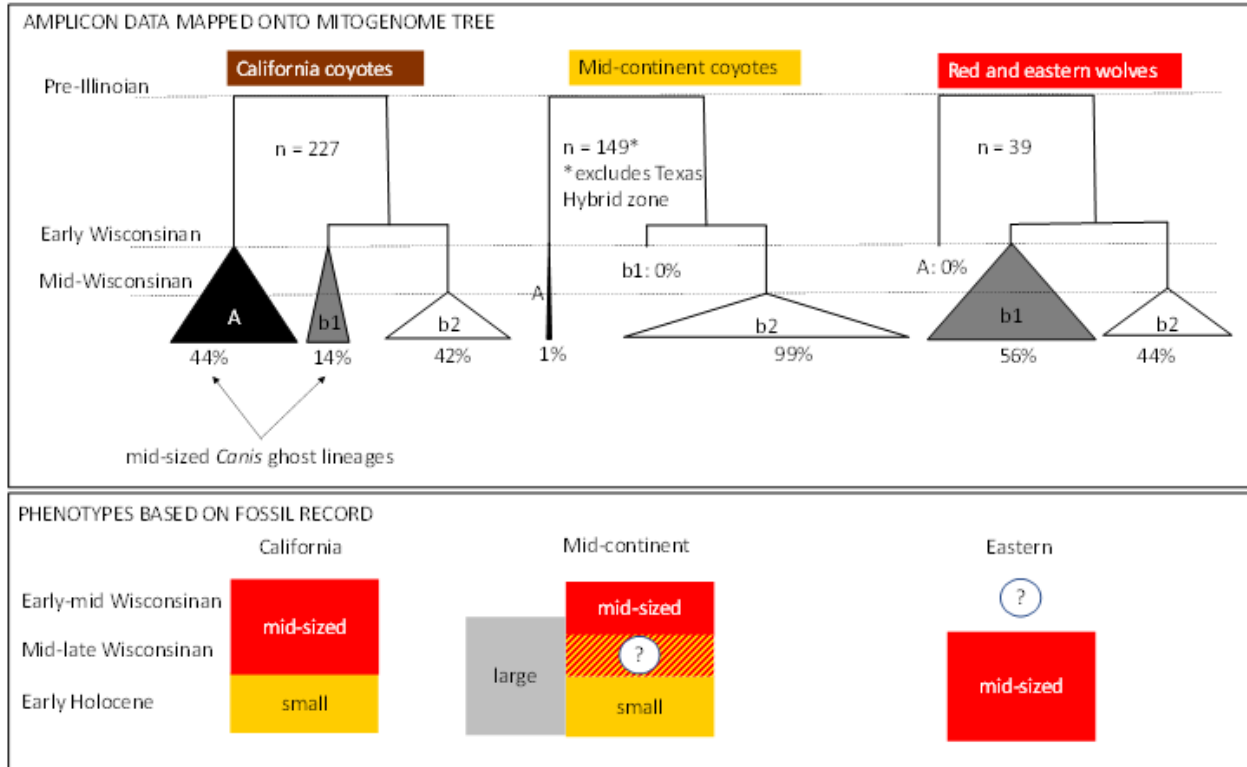
**Figure 1.** Sampling locations relative to estimated historical ranges (at the time of European settlement) of North American red wolves (*C. rufus*), eastern wolves (*C. lycaon*), coyotes (*C. latrans*), and gray wolves (*Canis lupus*): (A) 55 complete and one partial (\*) mitogenomes, including 11 historical specimens and 25 modern canids sequenced in this study, as well as 20 mitogenomes assembled from publicly available read datasets ( $n = 14$ ) or as direct downloads from GenBank ( $n = 6$  coyotes). Four modern red wolves descend from 12 ancestors captured in the 1970s in southeastern Texas or western Louisiana. Not shown is one Alaskan coyote used in this study; (B) 412 modern coyotes sequenced at mitochondrial cytochrome *b* (cytb) and D loop fragments (see Table S2); not shown are 4 cytb and D loop sequences from GenBank (Alaska, Mexico, Illinois, Missouri) and 9 samples sequenced only at the D loop region (Wyoming,  $n = 3$ ; Montana,  $n = 2$ ; Nebraska,  $n = 2$ ; Texas,  $n = 1$ ; Saskatchewan,  $n = 1$ ). Range maps were based on (Hody & Kays, 2018) for the coyote and (Chambers et al., 2012) and (Nowak, 2002) for the wolf taxa. Sample information is detailed in Tables S1, S2, and Figure S1.



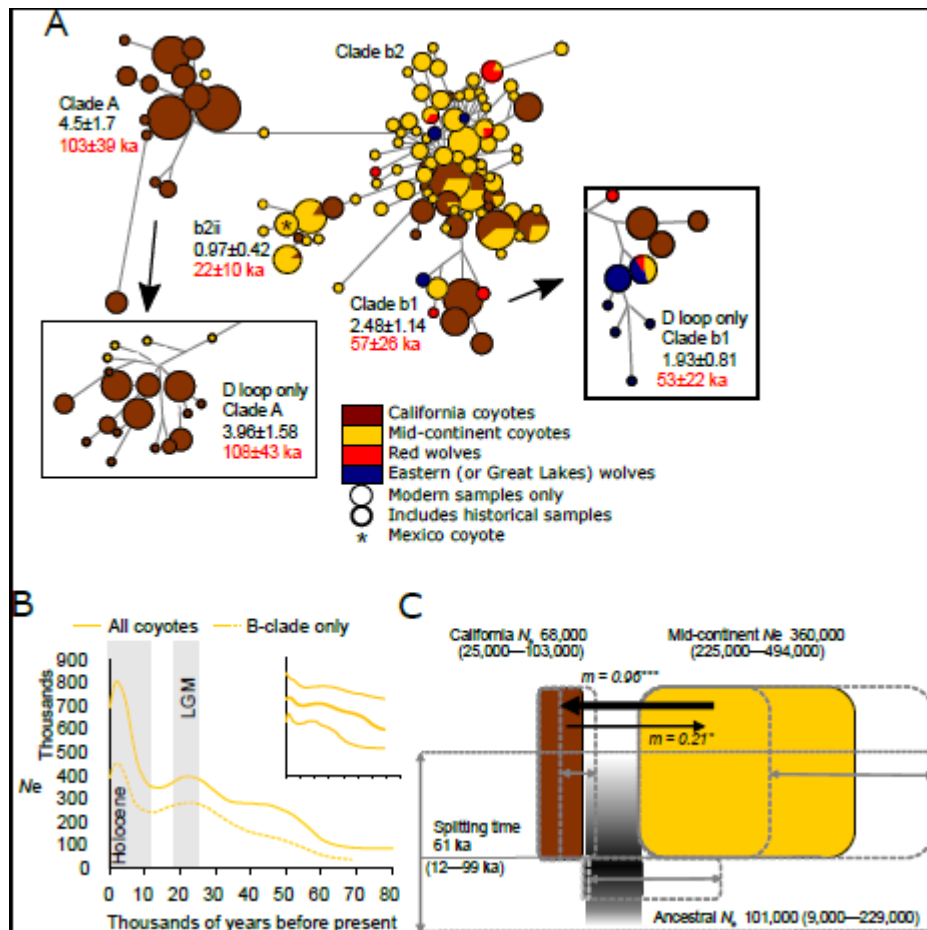
**Figure 2.** Analyses of 15,457 bp mitogenomes (control region excluded), including (A) phylogenetic tree constructed from third-codon positions of 155 *Canis* spp in Beast v 1.8.3 using a birth-death model with a relaxed lognormal clock and rooting to the African wild dog (*Lycaon pictus*), (B) expanded view of gray wolf clade, excluding basal Indian and Tibetan lineages, with Hokkaido specimens in pink, indicating the most recently divergent Eurasian (from North American) gray wolves (node a), (C) expanded view of American *Canis* clade, (D) chronology of selected node ages (95% HPD) as indicated on panels B and C and relative to the last glacial maximum (LGM) of the Wisconsinan glaciation and the Holocene, and (E) comparison of the nucleotide diversity of 13 red and eastern wolves (RW/EW) versus 16 mid-continent coyotes; (A-D) topology and node support were based on 15,457 bp of the mitogenome, but nodes were calibrated using only 3rd codon positions and a normal root height prior of 3.9 Ma (SD = 0.3 Ma) based on fossil calibrated nuclear DNA dating (Chavez et al., 2019); samples used in tree are mapped in Fig. 1A and described in the Supplemental Methods, Tables S1, S2, and Figure S1. See Supplemental Methods for partitions and mutation models.



**Figure 3.** Estimates of divergence time and historical demography of all coyotes (gold) and red (and eastern) wolves (red) based on mitogenomes (15,457 bp, excluding control region) of 36 coyotes and 12 red wolves (only 1 of 4 modern red wolf sequence was used in this analysis): (A) parameter estimates from an isolation-with-migration model including all 36 coyotes and 12 red wolves (full model) conducted in IMA2, (B) comparison of posterior distributions for splitting time among the full model, one restricted to the 18 mid-continent coyotes (i.e., excluding the 15 California and 3 eastern coyotes), and a third also excluding California and eastern coyotes and assuming no post-divergence gene flow, and (C) trajectories of effective population size ( $N_e$ ) over time indicated by Bayesian skyline plots (BSP) of red wolves and mid-continent coyotes. BSPs were estimated in Beast v 1.8.2 and Tracer v 1.6. (A) Parameters include  $N_e$ , splitting time, and migration ( $m$ ), equal in both directions; (C) Thick lines indicate median estimates of  $N_e$  in main plot and insets; thin lines in insets indicate 95% highest posterior density (HPD) traces and are shown on a log-scale; (B, C) Holocene epoch and last glacial maximum (LGM) are indicated for reference. Sequences used in analyses are presented in Table S2.



**Figure 4.** Proportional distribution of 415 amplicons across the three major American *Canis* clades as inferred from mitogenomes for California coyotes (left), mid-continent coyotes (center), and red and eastern wolves (right), illustrating shared ancestry of California coyotes and red and eastern wolves dating to the early Wisconsinan glaciation (clade b1) and more recent shared ancestry both between mid-continent coyotes and California coyotes and between mid-continent coyotes and red wolves dating to the mid-Wisconsinan glaciation (clade b2). Because they are essentially absent from mid-continent coyotes, clades A and b1 in California are presumed to reflect ghost lineages from the Pleistocene mid-sized *Canis* occurring in California prior to 12–10 ka, after which small *Canis* replaced them in the fossil assemblages. Below is a schematic of phenotype timelines evident in the fossil records of the three geographic partitions as described in the main text. Sample information and amplicon haplotypes are in Tables S3 and S4.



**Figure 5.** Analyses of combined sequences of the cytochrome *b* gene (369 bp) and D loop region (278 bp) of 227 California (brown) and 189 mid-continent (gold) coyotes: (A) median joining network illustrating phylogenetic relationships among coyotes relative to 7 historical and 8 modern red and eastern wolves, and insets showing D loop-only versions of two clades to accommodate published D loop samples; (B) trajectory of effective population sizes ( $N_e$ ) over time indicated by Bayesian skyline plots of all 416 coyote haplotypes and the 314 B-clade coyote haplotypes (i.e., excluding 98 A-clade haplotypes) conducted in BEAST v 1.8.2 and Tracer v 1.5, and (C) parameter estimates from an isolation-with-migration model between California and mid-continent coyotes conducted in IMA2. (A) Time estimates (red font) inferred from rho statistics (black font) (Macaulay et al., 2019) were based on calibration of the mutation rate to the mitogenomes, resulting in an estimate for the combined 647 bp sequence of 6.74%/Ma (1.9%/Ma and 13.2%/Ma for *cytb* and D loop, respectively, calibration described in methods). Insets show published D-loop sequences from (Fain et al., 2010) (Clade A, 1 modern mid-continent coyote) and (Leonard & Wayne, 2008) (subclade b1, 17 historical eastern wolves), illustrating shared ancestry between California coyotes and eastern wolf-like canids. The one mid-continent coyote haplotype in clade b1 is from the Texas hybrid zone. Entire D loop network is in Fig. S2. (B) Thick solid lines indicate median estimates of  $N_e$  in main plot and insets; thin lines in insets indicate 95% highest posterior density (HPD) traces and are shown on a log-scale; Holocene epoch and last glacial maximum (LGM) are indicated for reference; (C) IMA2 parameters include  $N_e$ , splitting time, and migration ( $m$ ), with arrows indicating direction of gene flow forward in time.