

1 Original Article

2 **Sequencing red fox Y chromosome fragments to develop phylogenetically informative SNP**
3 **markers and glimpse male-specific trans-Pacific phylogeography**

4

5 Sacks BN, Lounsberry ZL, Rando HM, Kluepfel K, Fain S, Brown SK, and Kukekova AV.

6 From the Mammalian Ecology and Conservation Unit of the Veterinary Genetics Laboratory,
7 University of California, Davis, Davis, California, USA (Sacks, Lounsberry, Kluepfel, Brown);
8 Department of Population Health and Reproduction, University of California, Davis, Davis,
9 California, USA (Sacks, Statham).

10 From the U. S. Fish & Wildlife Service, National Forensics Laboratory, Ashland, Oregon, USA
11 (Fain)

12 From the Department of Animal Sciences, College of Agricultural, Consumer and
13 Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (Rando,
14 Kukekova); Illinois Informatics Institute, University of Illinois at Urbana-Champaign, Urbana,
15 IL 61801 (Rando)

16 Address correspondence to Benjamin N. Sacks, Veterinary Genetics Laboratory, One Shields
17 Avenue/Old Davis Road, University of California, Davis, Davis, California 95616, USA or e-
18 mail: bnsacks@ucdavis.edu

19 **Running Title:** Red fox Y chromosome sequencing and SNP development

20 **Abstract**

21 The red fox (*Vulpes vulpes*) has a wide global distribution with many ecotypes, and has been
22 bred in captivity for various traits, making it a useful evolutionary model system. The Y
23 chromosome represents one of the most informative markers of phylogeography, yet it has not
24 been well-studied in the red fox due to a lack of the necessary genomic resources. We used a
25 target capture approach to sequence a portion of the red fox Y chromosome in a geographically
26 diverse red fox sample, along with other canid species, to develop single nucleotide
27 polymorphism (SNP) markers, 13 of which we validated for use in subsequent studies.
28 Phylogenetic analyses of the Y chromosome sequences, including calibration to outgroups,
29 confirmed previous estimates of the timing of two intercontinental exchanges of red foxes, the
30 initial colonization of North America from Eurasia approximately half a million years ago and a
31 subsequent continental exchange before the last Pleistocene glaciation (~100,000 years ago).
32 However, in contrast to mtDNA, which showed unidirectional transfer from Eurasia to North
33 America prior to the last glaciation, the Y chromosome appears to have been transferred from
34 North America to Eurasia during this period. Additional sampling is needed to confirm this
35 pattern and to further clarify red fox Y chromosome phylogeography.

36 **Key words:** target capture, *Vulpes fulva*, *V. vulpes*, Y chromosome

37

38 **Sequencing red fox Y chromosome fragments to develop SNP markers and glimpse male-**
39 **specific trans-Pacific phylogeography**

40

41 Sacks BN, Lounsberry ZL, Rando HM, Kluepfel K, Fain S, Brown SK, and Kukekova AV.

42

43 The phylogeography of red fox (*Vulpes vulpes*) has been well-characterized in terms of
44 mitochondrial and nuclear patterns of diversity over most of its global distribution (Aubry et al.
45 2009; Edwards et al. 2012; Kutschera et al. 2013; Statham et al. 2014, 2018; Goldsmith et al.
46 2016; Sacks et al. 2018). For example, the most divergent mitochondrial split dates to
47 approximately half a million years ago and separates a clade that evolved in North America
48 south of the ice sheets prior to the last glaciation (Nearctic clade) and one that spans Eurasia and
49 Alaska/western Canada, reflecting a secondary continental exchange event ~100 thousand years
50 ago (ky; Aubry et al. 2009; Statham et al. 2014). However, little is known about red fox Y
51 chromosome diversity. Except for microsatellites, no markers exist for such investigations in red
52 foxes (Statham et al. 2014; Rando et al. 2017). Although microsatellites are useful for
53 population-genetic questions about contemporary gene flow and recent historical demography,
54 more conserved mutations are needed for deeper reconstruction of Y chromosome phylogenies.

55

56 Sequencing the Y chromosome is challenging, even in model species, due to an evolutionarily
57 plastic history that has resulted in high frequency of repetitive DNA, paralogs of autosomal and
58 X chromosomal regions, and palindromes (Charlesworth and Charlesworth 2000; Skaletsky et al.
59 2003; Tomaszkiwicz et al. 2016). The challenge is even more pronounced in non-model
60 organisms due to a lack of reference genomes. Notwithstanding these obstacles, it is feasible to
61 enrich libraries for DNA resembling Y chromosome sequence from other mammals, sequence
62 these enriched libraries, and use read depth as a criterion for distinguishing unique loci from
63 repetitive motifs (Tomaszkiwicz et al. 2016; Rangavittal et al. 2018). As a final confirmation,
64 primers can be designed to genotype the presumptive unique loci and to test them in male and
65 female individuals (Natanaelsson et al. 2006). Those that turn out to be male-specific and
66 provide no more than one allele per individual can be inferred to be male-specific Y
67 chromosome markers. Because the entire male-specific length of the Y chromosome is linked,
68 these markers can be employed in tandem, regardless of their relative locations on the
69 chromosome, as multi-SNP haplotypes to reconstruct phylogenetic topologies.

70

71 In this study, we used information from the dog genome to enrich genomic fox libraries for Y
72 chromosome DNA and sequenced the libraries in a diverse geographic sample of red fox to
73 develop Y chromosome SNP markers and provide a preliminary look at male-specific
74 phylogeography. Although dogs and red foxes are relatively close relatives, they have distinct
75 karyotypes, and sex chromosomes in particular are evolutionarily unstable (Charlesworth and
76 Charlesworth 2000). Therefore, we emphasized use of Y chromosome regions that were
77 conserved across multiple orders of mammal (Natanaelsson et al. 2006) and validated SNP
78 markers on male and female red foxes. To obtain a red fox Y chromosome tree and divergence
79 time estimates for major Y chromosome clades, we conducted a phylogenetic analysis rooting
80 and calibrating branch lengths among red fox Y chromosome sequences, those of the most basal
81 extant canids (*Urocyon* spp.), and of coyotes (*Canis latrans*).

82

83 **Materials and Methods**

84 *Samples.*-- For high-throughput sequencing, we used 18 male red fox samples (Fig. 1), along
85 with 2 gray foxes (*Urocyon cinereoargenteus*), an island fox (*U. littoralis*), and 6 coyotes (*Canis*
86 *latrans*) as outgroups (Supplementary Table S1). These samples were used for SNP discovery
87 and phylogenetic analyses. We used 9 of the same red fox males, one additional red fox male
88 from Alaska, and 19 red fox females for red fox SNP marker validation.

89 *Capture enrichment and sequencing.*--We constructed individually barcoded whole genome
90 shotgun libraries using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England
91 Biolabs, Ipswich, MA) and 6-bp custom indexed primers. We then pooled libraries (10 ng each)
92 in batches of 47 samples for hybridization to capture baits. Although the entire dog Y
93 chromosome assembly is now available (Li et al. 2013), we began this study just prior to its
94 publication and, therefore, relied on a smaller subset of canine Y chromosome fragments totaling
95 39,246 bp. (However, we did map markers to the Li et al. 2013; GenBank Accession No.
96 KP081776.1.) We designed baits that were complementary to previously published and
97 validated dog Y-chromosome DNA sequences (Supplementary Information 1) that had been
98 prescreened to avoid paralogy to X chromosome sequences from canids and other mammals
99 (Hellborg and Ellegren 2003; Natanaelsson et al. 2006; Ding et al. 2012; Tsubouchi et al. 2012).
100 We also included capture baits for exons from the Y chromosome amelogenin (AMELY) and
101 zinc finger (ZFY) genes, which do have high paralogy with the X chromosome, but could have
102 unique regions as well. Baits were each 80 bp in length and overlapped adjacent baits by 40 bps
103 (i.e., tiling density = 2x), as appropriate for capture between closely related species. The RNA
104 baits were synthesized and biotinylated by MyBaits (MYcroarray, Ann Arbor, MI). We
105 conducted the hybridization with blockers for 36 hours at 65°C according to manufacturer's
106 protocols, using the MyBaits Target Enrichment Kit (Arbor Biosciences, Ann Arbor, MI, USA)
107 We bound hybridization products to streptavidin-coated magnetic beads, washed away unbound
108 library, isolated bound (i.e., enriched) library from baits and beads, and PCR-amplified the
109 enriched library for 14 cycles. We then sequenced the pooled enriched library on an Illumina
110 MiSeq lane SR300 at the UC Davis Genome Center. Additionally, we included HiSeq2500
111 reads (PE 150) from 2 male red fox whole genome shotgun libraries sequenced for a different
112 project.

113 *Bioinformatic processing.*— We trimmed adapter sequences and filtered reads for quality,
114 discarding reads with >50% of bases with quality scores <2, and trimmed low-quality 3' end
115 bases using Next Generation Sequencing Short Read Trimmer (ngsShoRT; Chen *et al.* 2014).
116 We aligned reads to the original bait sequences using bwa-mem (Li and Durbin 2009) and
117 processed alignments with Samtools (Li et al. 2009). We then used freeBayes v0.9.21 (Garrison
118 and Marth 2012) to call variants against the original bait sequence references, allowing
119 “heterozygotes” to flag paralogs downstream, and exported all positions for each individual
120 (variable and not) in variable call format (vcf). Because we had no independent knowledge of
121 X-degenerate (uniquely mapping) red fox or *Urocyon* Y chromosome sequence, we used
122 coverage depth as a basis for selecting positions empirically. The distribution of coverage depths
123 within a species contained multiple modes, which we presumed corresponded on the low end to
124 erroneous sites (e.g., exogenous contamination), followed by uniquely aligning sites, sites
125 aligning to both sex chromosomes, and, lastly, a long tail corresponding to variable frequency

126 repetitive motifs. Therefore, after visualizing the distributions for red foxes, we selected a
127 coverage-depth range corresponding to the second mode and either tail (see Results). We used
128 these sites to obtain intraspecific SNPs.

129 *Phylogenetic analysis.*—We further filtered the sites inferred above to be uniquely aligning to
130 the red fox Y chromosome based on the same criteria described above. We also removed any
131 interspecific indels to obtain Y chromosome sequence orthologous across all taxa for calibrating
132 substitution rates against independently determined node ages (Wayne et al. 1997; Lindblad-
133 Toh et al. 2005; Perini et al. 2009). We used Mega 6 (Tamura et al. 2013) to construct a
134 maximum likelihood tree with 500 bootstrap replicates based on the Tamura 3 parameter
135 mutation model (T92; Tamura 1992), which was supported over 23 other models by the lowest
136 Bayesian Information Criterion. To estimate divergence time between red fox clades, we
137 computed pairwise T92 distances and standard errors between red fox clades and calibrated these
138 to pairwise distances to coyote and gray fox, previously estimated to be 8.8 MY divergent from
139 red foxes (Perini et al. 2009).

140 *SNP assay.*—We used Assay Design Suite 2.0 (<https://mysequenom.com/Tools>) to design
141 primers for multiplexing on the MassARRAY iPLEX platform (Agena Biosciences, Inc., San
142 Diego, CA). To validate these SNPs as uniquely representing the non-recombining portion of the
143 Y chromosome, we tested them in both male and female foxes and retained only those yielding
144 male-specific genotypes.

145 *Intraspecific red fox network.*—After combining sequencing and genotyping calls, a small
146 number of remaining missing calls were imputed parsimoniously with the allele found in
147 otherwise identical complete sequences. We constructed statistical parsimony network using the
148 median joining method in Networks (v 5.0; Fluxus Technology Ltd, England; Bandelt et al.
149 1999; Forster et al. 2000) and estimated the age of clades in terms of the average (and standard
150 error) numbers of mutations separating derived from inferred ancestral nodes (Forster et al. 2000;
151 Saillard et al. 2000).

152 **Results**

153 Excluding one red fox sample that failed sequencing, we obtained 6,486,865 raw 300-bp reads
154 from a MiSeq lane of the 23 target-enriched samples of which 2,393,840 (36.8%) reads aligned
155 back to the original bait sequences (i.e., 41 targeted fragments of the dog Y chromosome). This
156 total included 1,342,241 of 3,304,125 reads (40.6%) from 14 red foxes, 738,554 of 2,296,795
157 reads (32.2%) from 6 coyotes, and 313,045 of 898,307 reads (34.8%) from 3 *Urocyon* foxes. We
158 also obtained 86,687,990 bp of HiSeq sequence for 2 additional red foxes, of which 1,835,416 bp
159 (2.1%) aligned to the bait sequences. The ratio of targeted sequence obtained from the capture
160 enrichment vs whole genome shotgun approaches for red foxes (i.e., 40.6% vs 2.1%) indicate
161 that the enrichment process increased yield approximately 20-fold over random sequencing.

162 In total, we recovered 37,080 sites, of which 31,624 (85.3%) had coverage depths in the range
163 we presumed corresponded to X-degenerate sites on the Y chromosome (10–160×) and no
164 heterozygous sites in red foxes (Fig. 2; vcf file deposited in Dryad). For phylogenetic analyses,
165 we retained 20,931 of these sites after removing interspecific indels and confirming that read
166 depths in the other species were also consistent with uniquely aligning sites. These 20,931 sites

167 had an average of 85× depth in red foxes (5.3× per individual), 85× in coyotes (14.3× per
168 individual), and 30× in *Urocyon* spp. (10× per individual).

169 In the 20,931 site dataset (all taxa), we observed 460 substitutions, including 54 intraspecific (or
170 intrageneric in the case of *Urocyon*) polymorphisms and 404 interspecific substitutions (254 Ti,
171 150 Tv). Two of the 54 intraspecific polymorphisms occurred in both red foxes and *Urocyon*,
172 altogether yielding 22 polymorphisms (14 Ti, 8 Tv) in 17 red foxes, 13 polymorphisms (9 Ti, 4
173 Tv) among the 3 *Urocyon* foxes, and 23 polymorphisms (14 Ti, 9 Tv) among the 6 coyotes. A
174 maximum likelihood tree indicated two well-supported, reciprocally monophyletic red fox clades
175 (Fig. 3). Based on the Tamura 3 parameter model distances (and SE) calibrated to *Urocyon* and
176 coyote divergence times (8.8 MY, Perini et al. 2009), the two red fox Y chromosome clades
177 diverged approximately 470,000 (402,000–563,000) years ago, which was consistent with
178 mitochondrial and autosomal estimates (Statham et al. 2014).

179 For intraspecific analyses of red foxes, we retained all of the 31,624 presumptive Y chromosome
180 sites. Among these sites, we observed 31 SNPs within red foxes (including the 22 intraspecific
181 sites used above, along with 9 additional sites); we designed SNP assays for 14 of them (Table 1;
182 Supplementary Tables S2, S3). We tested SNP assays on 9 of the originally sequenced male red
183 foxes and 19 female red foxes and also genotyped the Alaskan male that failed sequencing and
184 one other Alaskan male for which no sequencing was attempted. All genotypes for the 9 males
185 previously sequenced matched those based on the original sequences. All except one of the loci
186 failed to amplify consistently in females. Locus 20_357AT amplified in 14 of 19 female
187 samples, indicating that the primers we designed were not sufficiently specific to the Y
188 chromosome target. Otherwise, the false-positive rate among the other 13 loci was 3.5% in
189 females, with no single locus or individual yielding notably more than any other. Positions of
190 the 13 validated SNPs were determined based on the Y chromosome dog assembly
191 (Supplementary Table S3).

192 We constructed two median joining networks: one using all 31 variable sites in the 17
193 successfully sequenced red foxes (Fig. 3A), and one using the subset of 13 assayed and validated
194 sites that additionally included 2 Alaskan male red foxes genotyped using the assay (Fig. 3B;
195 Table 1). The former network provided more resolution of haplotypes, particularly within
196 continents, whereas the latter network included two additional Alaskan foxes with haplotypes
197 that were basal within their clade. For the network based on all 31 sites, the rho estimates (i.e.,
198 average numbers of mutations since divergence from the most recent common ancestor) were
199 13.00 (SD = 2.55) for the entire network and 3.82 (SD = 1.48) for the North American/Siberian
200 clade, indicating a ratio of 0.294. For the network based on 13 sites, the rho estimates were 4.80
201 (SD = 1.55) for the entire network and 1.77 (SD = 0.98) for the North American/Siberian clade,
202 indicating a ratio of 0.369. Assuming a 470,000 year divergence between the two clades, these
203 estimates imply an age of the most recent common ancestor of Siberian and North American red
204 foxes ranging 77,000 to 270,000 years ago based on the 13-site network or 85,000 to 190,000
205 years ago based on the 31-site network, both of which are consistent with previous mtDNA and
206 autosomal nuclear estimates (Statham et al. 2014).

207 Discussion

208 The Y chromosome is replete with repetitive elements and regions paralogous to other
209 chromosomes, posing several significant obstacles to sequencing even in species for which the Y

210 chromosome genome has been assembled and provides a reference (Tomaszkiewicz et al. 2016).
211 Our reliance on dog Y chromosome sequence, which likely shared only partial homology to red
212 and gray foxes, complicated this task further. Nevertheless, our efforts to enrich fox libraries for
213 Y chromosome reads and obtain sequence for non-repetitive fox Y chromosome fragments were
214 successful.

215 Overall, we obtained approximately 75% of the targeted Y chromosome sequence as uniquely
216 aligning in male red foxes. Although 1 of the 14 SNPs tested in females yielded regular
217 amplification products, this may have been due to the non-specificity of the primers themselves
218 rather than of the contiguous sequence reflected in the sequencing reads. The read depth at this
219 SNP in the original sequencing reads was 132× among red foxes, which was well within the
220 range of depths observed in most positions and substantially smaller than the range
221 corresponding to a larger mode and presumed to correspond to loci with X and Y paralogs.
222 Regardless, 13 of the loci for which we developed a genotyping assay resolved two major clades,
223 as well as several haplotypes within each. We also discovered 18 additional polymorphisms for
224 which SNP assays could be developed and validated in the future. Most of these 18 SNPs would
225 further differentiate the already well-defined major clades (Table 1). More importantly, 3
226 intraspecific polymorphisms (i.e., 29_231AC, MS41A_209CT, MS41A_248AG) may be useful
227 in further resolving North American haplotypes.

228 Given the nearly global distribution of red foxes and many translocations (Long 2003), these
229 markers will be useful in complementing mitochondrial markers as a way of easily identifying
230 continent of origin, for example, in the case of introductions (e.g., Kasproicz et al. 2016).
231 Moreover, combining these SNPs with Y-linked microsatellites (Rando et al. 2017) can elucidate
232 historical demographic patterns on Holocene time scales (Forster et al. 2000; Sacks et al. 2013).

233 Although our study was not designed primarily as a phylogeography study, which would require
234 many more samples, our findings nevertheless provided some insights about the red fox's
235 continental Y chromosome phylogeography in the context of previous models of continental
236 exchange between Eurasian and North American red foxes. Most fundamentally, our Y
237 chromosome phylogeny provided independent support for previous estimates of the timing of
238 continental exchanges between red foxes, specifically, an initial colonization of North America
239 by Eurasian red foxes approximately 500 ky and a secondary continental exchange around the
240 beginning of the last glaciation ~100 ky (Aubry et al. 2009; Statham et al. 2014).

241 Prior to this study, discordance observed between mitochondrial and nuclear genetic patterns
242 resulted in unresolved hypotheses about the nature and extent of the secondary exchange
243 (Statham et al. 2014; Sacks et al. 2018). In particular, nearly all of the mitochondrial haplotypes
244 observed in modern red foxes from Alaska and much of northwest Canada traced to Eurasia
245 around the time of the secondary exchange (e.g., Aubry et al. 2009; Statham et al. 2014;
246 Goldsmith et al. 2016). In contrast, most of the nuclear genetic ancestry in this same region
247 traced to the original colonization of North America 500 ky, with little genetic exchange in either
248 direction corresponding to the secondary contact event (Statham et al. 2014; Sacks et al. 2018).
249 These discordant patterns suggested two hypotheses: (1) a selective sweep on Eurasian mtDNA
250 introduced into North American red foxes following the secondary exchange, or (2) a male-

251 mediated expansion of southern North American red foxes after the recession of the two major
252 North American ice sheets after the last glacial maximum (<20 ky), leading to partial
253 replacement of a Beringian population formerly composed of Eurasian red foxes. These
254 hypotheses differ in whether the pre-event (continental exchange, expansion) population in
255 Beringia was composed of Eurasian or North American red foxes and in the timing of the event
256 that led to the discordance.

257 In the present study, we found the major split in the Y chromosome phylogeny dated to 500 ky,
258 presumably reflecting the same secondary intercontinental split seen in nuclear and
259 mitochondrial data. However, one of the clades included all North American, as well as
260 Siberian, samples. The estimated age of the most recent common ancestral haplotype of Siberian
261 and North American haplotypes dated to approximately 100 ky, consistent with Y chromosome
262 movement from North America to Eurasia during the intercontinental exchange. This
263 observation is consistent with hypothesis 1 but not hypothesis 2, which implies North American
264 patrines did not occur in Beringia until after the second continental exchange ~100 ky. Thus,
265 the exclusivity of Eurasian mitochondrial haplotypes combined with the general rarity of
266 Eurasian nuclear ancestry in modern-day Alaskan red foxes most likely traces to a selective
267 sweep ~100 ky on the mitochondrial genome.

268 Clearly our sample size was too small to address the extent to which North American Y
269 chromosomes contributed to East Asian red fox diversity, or whether some reciprocal exchange
270 could have occurred. These questions can be investigated further in the future by genotyping
271 larger samples of foxes with the SNPs developed in this study, and in combination with Y
272 chromosome microsatellite markers (Statham et al. 2014; Rando et al. 2017).

273 **Supplementary Material**

274 Supplementary data are available at Journal of Heredity online.

275 **Funding**

276 This work was supported by the Mammalian Ecology and Conservation Unit and the UC Davis
277 Forensic Sciences Graduate Program.

278 **Acknowledgments**

279 We thank H Liu for assistance in sample preparation. We thank B. Popper, B. Chomel, C.
280 Soulsberry, C. Miller, D. Berteaux, D. Ehrich, E. Tiller, J. Velasquez, J. Amaral, J. Perrine, J.
281 Lindeman, M. Reid, P. Magee, P. Cross, P. Cervantez, S. Lariviere, T. Brinkman, and W.
282 Carleson for red fox samples.

283 **Data Availability**

284 We have deposited the primary data underlying these analyses as follows:

- 285 - Y chromosome variable call format (.vcf) file: Dryad
- 286 - Sample information, primers, and flanking sequence uploaded as online Supplementary
- 287 Material

288 **Supplementary Material**

289 Supplementary Information 1

290 -doc file

291 Supplementary Tables S1–S3

292 -xls file

293 References

- 294 Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM. 2009. Phylogeography of the North
295 American red fox: vicariance in Pleistocene forest refugia. *Mol Ecol.* 18:2668–2686.
- 296 Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific
297 phylogenies. *Mol Biol Evol.* 16:37–48.
- 298 Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R*
299 *Soc Lond B Biol Sci* 355:1563–1572.
- 300 Chen C, Khaleel SS, Huang H, Wu CH. 2014. Software for pre-processing Illumina next-
301 generation sequencing short read sequences. *Source Code Biol Med.* 9:8.
- 302 Ding ZL, Oskarsson M, Ardalán A, Angleby H, Dahlgren LG, Tepeli C, Kirkness E, Savolainen
303 P, Zhang YP. Origins of domestic dog in southern East Asia is supported by analysis of
304 Y-chromosome DNA. *Hered.* 108:507–514.
- 305 Edwards CJ, Soulsbury CD, Statham MJ, Ho SY, Wall D, Dolf G, Iossa G, Baker PJ, Harris S,
306 Sacks BN, Bradley DG. 2012. Temporal genetic variation of the red fox, *Vulpes vulpes*,
307 across western Europe and the British Isles. *Q Sci Rev.* 57:95–104.
- 308 Forster P, Röhl A, Lünemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A
309 short tandem repeat–based phylogeny for the human Y chromosome. *Am J Hum Genet.*
310 67:182–196.
- 311 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.
312 *arXiv preprint arXiv:1207.3907* [q-bio.GN]
- 313 Goldsmith EW, Renshaw B, Clement CJ, Himschoot EA, Hundertmark KJ, Hueffer K. 2016.
314 Population structure of two rabies hosts relative to the known distribution of rabies virus
315 variants in Alaska. *Mol Ecol.* 25:675–688.
- 316 Hellborg L, Ellegren H. 2003. Y chromosome conserved anchored tagged sequences (YCATS)
317 for the analysis of mammalian male-specific DNA. *Mol Ecol.* 12:283–291.
- 318 Kasprovicz AE, Statham MJ, Sacks BN. 2016. The fate of the other red coat: remnants of
319 colonial British red foxes in the Eastern United States. *J Mammal.* 97:298–309.
- 320 Kutschera VE, Lecomte N, Janke A, Selva N, Sokolov AA, Haun T, Steyer K, Nowak C, Hailer
321 F. 2013. A range-wide synthesis and timeline for phylogeographic events in the red fox
322 (*Vulpes vulpes*). *BMC Evol Biol.* 13:114.
- 323 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
324 *Bioinform.* 25:1754–1760.
- 325 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R
326 and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence
327 alignment/map (SAM) format and SAMtools. *Bioinform.* 25:2078–2079.
- 328 Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. 2005.
329 Genome sequence, comparative analysis and haplotype structure of the domestic dog.
330 *Nature.* 438:803–819.

- 331 Long, JL. 2003. Introduced mammals of the world: their history, distribution, and influence,
332 CSIRO Publishing, Collingwood, Victoria, Australia.
- 333 Natanaelsson C, Oskarsson MC, Angleby H, Lundeberg J, Kirkness E, Savolainen P. 2006. Dog
334 Y chromosomal DNA sequence: identification, sequencing and SNP discovery. *BMC*
335 *Genet.* 7:45.
- 336 Perini FA, Russo CAM, Schrago CG. 2010. The evolution of South American endemic canids: a
337 history of rapid diversification and morphological parallelism. *J Evol Biol.* 23:311–322.
- 338 Rando H, Stutchmann JT, Bastounes ER, Johnson JL, Driscoll CA, Trut L, Sacks BN, Kukekova
339 AV. 2017. Y-chromosome markers for the red fox (*Vulpes vulpes*), *J Hered.* 108:678–685.
- 340 Rangavittal S, Harris RS, Cechova M, Tomaszkiwicz M, Chikhi R, Makova KD, Medvedev P.
341 2017. RecoverY: K-mer-based read classification for Y-chromosome-specific sequencing
342 and assembly. *Bioinform.* 34:1125–1131.
- 343 Sacks BN, Stephens D, Brown SK, Pedersen NC, Wu J-T, Berry O. 2013. Y chromosome
344 analysis of dingoes and Southeast Asian village dogs suggests a continental Neolithic
345 expansion from Southeast Asia followed by multiple Austronesian dispersals. *Mol Biol*
346 *Evol.* 30:1103–1118.
- 347 Sacks BN, Lounsbury ZL, Statham MJ. 2018. Nuclear genetic analysis of the red fox across its
348 trans-Pacific range. *J Hered.* 109:573–584.
- 349 Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. 2000. mtDNA variation among
350 Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet.* 67:718–726.
- 351 Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S,
352 Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y
353 chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- 354 Statham MJ, Murdoch J, Janecka J, Aubry KB, Edwards CJ, Soulsbury CD, Berry O, Wang Z,
355 Harrison D, Pearch M, Tomsett L, Chupasko J, Sacks BN. 2014. Range-wide multilocus
356 phylogeography of the red fox reveals ancient continental divergence, minimal genomic
357 exchange, and distinct demographic histories. *Mol Ecol.* 23:4813–4830.
- 358 Statham MJ, Edwards CJ, Noren K, Soulsbury CD, Sacks BN. 2018. Genetic analyses of
359 European red foxes reveals multiple distinct peripheral populations and central
360 continental admixture. *Q Sci Rev.* 197:257–266.
- 361 Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong
362 transition-transversion and G + C-content biases. *Mol Biol Evol.* 9:678–687.
- 363 Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. 2013. MEGA6: molecular evolutionary
364 genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
- 365 Tomaszkiwicz M, Rangavittal S, Cechova M, Sanchez RC, Fescemyer HW, Harris R, Ye D,
366 O'Brien PC, Chikhi R, Ryder OA, Ferguson-Smith MA. 2016. A time- and cost-effective
367 strategy to sequence mammalian Y Chromosomes: an application to the de novo
368 assembly of gorilla Y. *Genome Res.* 2016 Mar 2.
- 369 Tsubouchi A, Fukui D, Ueda M, Tada K, Toyoshima S, Takami K, Tsujimoto T, Uruguchi K,
370 Raichev E, Kaneko Y, Tsunoda H, Masuda R. 2012. Comparative molecular
371 phylogeny and evolution of sex chromosome DNA sequences in the Family Canidae
372 (Mammalia: Carnivora). *Zool Sci.* 29:151–161.
- 373 Wayne RK, Geffen E, Girman DJ, Koepfli KP, Lau LM, Marshall CR. 1997. Molecular
374 systematics of the Canidae. *Syst Biol.* 46:622–653.
- 375

376 **Figure captions**

377 Figure 1. Locations of male red fox samples used in this study (n = 19). One sample from Alaska was
378 genotyped for selected SNPs but not sequenced. Numbers refer to the sample size indicated by a
379 marker. Markers with no numbers are a single sample.

380 Figure 2. Coverage depths from sequencing reads of 19 male red foxes mapped to 37,080 sites across
381 41 orthologous dog Y chromosome fragments shown (A) with read depths arrayed across ordered sites
382 on the reference fragments, and (B) as a frequency distribution, illustrating a systematic pattern of
383 coverage presumably corresponding to uniquely mapping Y chromosome sites (10-160x), sites likely to
384 reflect 2 paralogs (e.g., including one on the X chromosome, 161-210x), and sites with >2 paralogs
385 (>210x). Only sites with depths ranging 10-160x were used in analyses. Sites with depths <10x were
386 presumed to reflect errors and those >160x were presumed to reflect paralogs, not necessarily all on the
387 Y chromosome.

388 Figure 3. Red fox Y chromosome phylogeny displayed as Maximum Likelihood tree inferred from 20,931
389 bp of sequence in 17 red foxes, 2 gray foxes, an island fox, and 6 coyotes with node support inferred
390 from 500 bootstrapped trees. Genetic distance was computed according to the Tamura 3-parameter
391 model (Tamura 1992) and calibrated to divergence time in millions of years (MY).

392 Figure 4. Median joining networks utilizing (A) 31 Y chromosome SNPs in 17 red foxes or (B) a subset of
393 13 assayed Y chromosome SNPs in 19 red foxes (includes 2 additional Alaskan red foxes); the ancestral
394 node (or root, white circle) in each network corresponded to the haplotype common to all coyotes, gray
395 foxes, and the island fox. Red dots in (A) represent inferred nodes. (A,B) Nodes (observed or inferred)
396 used for aging haplogroups through rho statistics are indicated with an asterisk. Node sizes are
397 proportional to number of foxes and branch lengths are proportional to the number of mutations.

398

Table 1. Y chromosome haplotypes of 19 male red foxes and an ancestral node inferred from orthologous positions in *Canis* and *Urocyon*, which are composed of 31 variable sites discovered 31,624 bp of Y chromosome sequence. We also designed a genotyping assay for 13 of the loci (first 13 from left to right). Eighteen male red foxes were sequenced, including 9 foxes that were also genotyped (*), and 1 fox was genotyped but not sequenced (**). One locus (20_349AT) that amplified in 14 of 19 females was excluded from this table. None of the other 13 genotyped loci consistently amplified in females, although we observed a low (3.5%) false-positive rate. Genotype and sequencing calls matched 100% of the time. Two foxes that were genotyped but not sequenced have missing data (X) across all non-assayed sites.

Population	Sample identifier	Single nucleotide polymorphisms of 31 Y chromosome loci ^a
Ancestral node		G C C C C T T C T A G C A A G T T T G C A C T C A C T T A C G
Britain	S08-0428	C T - - - C - - T - - - A - - A - A C A C - - - - - - - -
Britain	S08-0433	C T - - - C - - T - - - A - - A - A C A C - - - - - - - -
Britain	S08-0436	C T - - - C - - T - - - A - - A - A C A C - - - - - - - -
Iraq	S10-0036*	C - - - - C - - T - - - G A C - A - A C A C - - T - - - T A
Iraq	S10-0039	C - - - - C - - T - - - G A C - A - A C A C - - T - - - T A
Iraq	S14-1369	C - - - - C - - T - - - G A C - A - A C A C - - - - - T A
Yamal, Siberia	S12-0237*	- - A T T C - T G - - - - - - - - A - C - - - - T G G G - -
Yamal, Siberia	S12-0241*	- - A T T C - T G - - - - - - - - A - C - - - - T G G G - -
Yamal, Siberia	S12-0244*	- - A T T C - T G - - - - - - - - A - C - - - - T G G G - -
Alaska, USA	S12-1162	- - A T T C - - G - C T - - - - C - A - C - - T - T G G G - -
Bylott Island, Canada	S12-1600*	- - A T T C - - G - C T T - - - C - A - C - - T G T G G G - -
Eastern Canada	S12-1607*	- - A T T C - - G - C T - - - - C - A - C - - T - T G G G - -
Eastern Canada	S12-1608*	- - A T T C - - G - C T - - - - C - A - C - - - - T G G G - -
Eastern Canada	S14-0418	- - A T T C - - G - C T T - - - C - A - - - - T G T G G G - -
Eastern Canada	S14-0432	- - A T T C - - G - C T T - - - C - A - - - - T G T G G G - -
Lassen Co, CA, USA	M1	- - A T T C - - G - C T - - - - C - A - C - - - - T G G G - -
Gunnison, CO, USA	S12-1176*	- - A T T C - - G - C T - - - - C - A - C - - T G T G G G - -
Alaska, USA	S12-1163*	- - A T T C - - G - - - - X X X X X X X X X X X X X X X X X X
Alaska, USA	S12-1161**	- - A T T C - - G - C - - X X X X X X X X X X X X X X X X X X

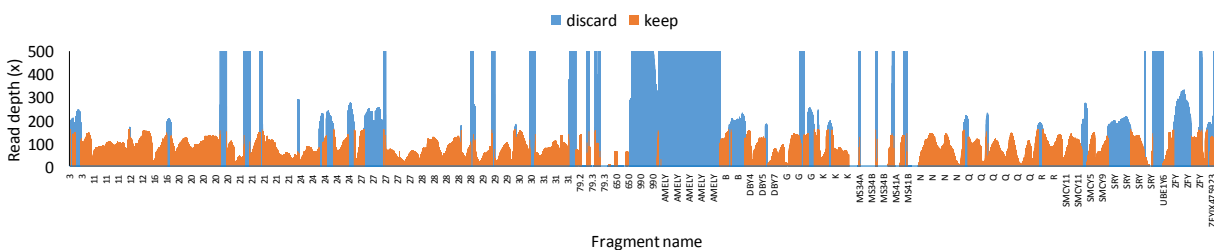
^aFrom left to right are the assayed loci (G_343GC, G_690CT, R_156AC, R_503CT, R_793CT, 12_709CT, 20_238CT, 21_687CT, 27_204GT, 31_365AT, DBY4_-54CG, SRY_1197CT, 31_209AT), followed by the un-assayed loci (11_535AG, 21_982GA, 21_1014TC, 24_878TC, 24_1434TA, 28_458GA, 28_524CA, 29_231AC, K_79CA, K_527TC, MS41A_209CT, MS41A_248AG, MS41B_221CT, N_558TG, Q_311TG, Q_322AG, Q_613CT, R_805G).

Figures



Figure 1. Locations of male red fox samples used in this study ($n = 19$). One sample from Alaska was genotyped for selected SNPs but not sequenced. Numbers refer to the sample size indicated by a marker. Markers with no numbers are a single sample.

A.



B.

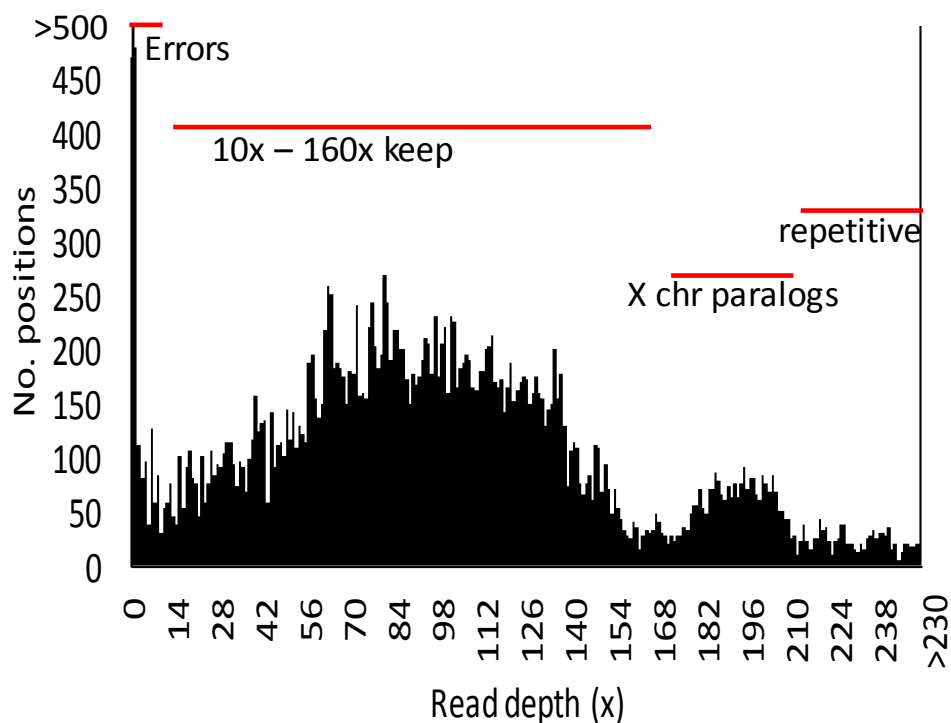


Figure 2. Coverage depths from sequencing reads of 19 red foxes mapped to 37,080 sites of 41 orthologous dog Y chromosome fragments shown (A) with read depths arrayed across ordered sites on the reference fragments, and (B) as a frequency distribution, illustrating a systematic pattern of coverage presumably corresponding to uniquely mapping Y chromosome sites (10-160x), sites likely to reflect 2 paralogs (e.g., including one on the X chromosome, 161-210x), and sites with >2 paralogs (>210x). Only sites with depths ranging 10-160x were used in analyses. Sites with depths <10x were presumed to reflect errors and those >160x were presumed to reflect paralogs, not necessarily all on the Y chromosome.

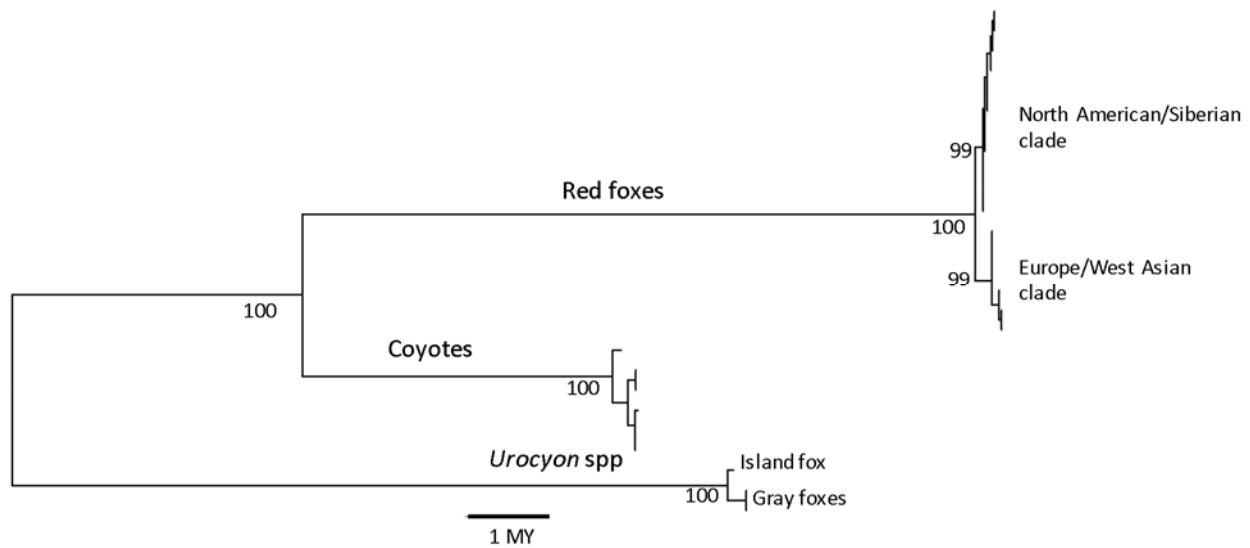


Figure 3. Red fox Y chromosome phylogeny displayed as Maximum Likelihood tree inferred from 20,931 bp of sequence in 17 red foxes, 2 gray foxes, an island fox, and 6 coyotes with node support inferred from 500 bootstrapped trees. Genetic distance was computed according to the Tamura 3-parameter model (Tamura 1992) and calibrated to divergence time in millions of years (MY).

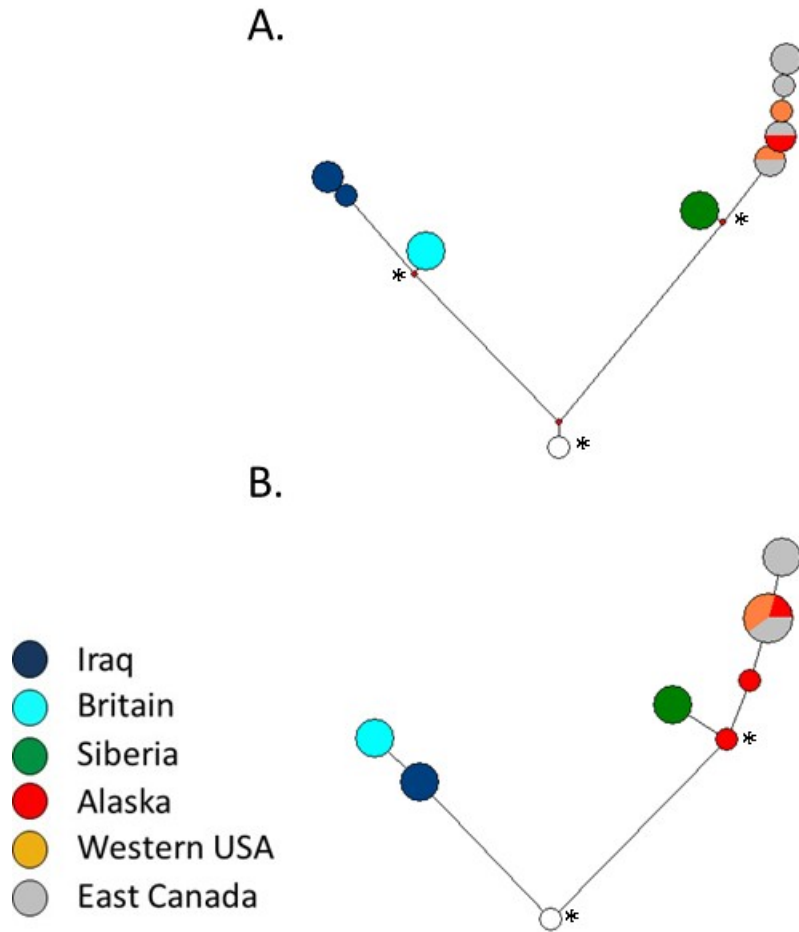


Figure 4. Median joining networks utilizing (A) 31 Y chromosome SNPs in 17 red foxes or (B) a subset of 13 assayed Y chromosome SNPs in 19 red foxes (includes 2 additional Alaskan red foxes); the ancestral node (or root, white circle) in each network corresponded to the haplotype common to all coyotes, gray foxes, and the island fox. Red dots in (A) represent inferred nodes. (A,B) Nodes (observed or inferred) used for aging haplogroups through rho statistics are indicated with an asterisk. Node sizes are proportional to number of foxes and branch lengths are proportional to the number of mutations.