

TECHNICAL ADVANCES

Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae

BENJAMIN N. SACKS* and SUSAN LOUIE†

**Canid Diversity and Conservation Project, Veterinary Genetics Laboratory, School of Veterinary Medicine, University of California, Davis, Davis, CA 95616-8744, USA*, †*College of Biological Sciences, University of California, Davis, Davis, CA 95616-8744, USA*

Abstract

Single nucleotide polymorphisms (SNP) are the ideal marker for characterizing genomic variation but can be difficult to find in nonmodel species. We explored the usefulness of the dog genome for finding SNPs in distantly related nonmodel canids and evaluated so-ascertained SNPs. Using 40 primer pairs designed from randomly selected bacterial artificial chromosome clones from the dog genome, we successfully sequenced 80–88% of loci in a coyote (*Canis latrans*), grey fox (*Urocyon cinereoargenteus*), and red fox (*Vulpes vulpes*), which compared favourably to a 60% success rate for each species using 10 primer pairs conserved across mammals. Loci were minimally heterogeneous with respect to SNP density, which was similar, overall, in a discovery panel of nine red foxes to that previously reported for a panel of eight wolves (*Canis lupus*). Additionally, individual heterozygosity was similar across the three canids in this study. However, the proportion of SNP sites shared with the dog decreased with phylogenetic divergence, with no SNPs shared between red foxes and dogs. Density of interspecific SNPs increased approximately linearly with divergence time between species. Using red foxes from three populations, we estimated F_{ST} based on each of 42 SNPs and 14 microsatellites and simulated null distributions conditioned on each marker type. Relative to SNPs, microsatellites systematically underestimated F_{ST} and produced biased null distributions, indicating that SNPs are superior markers for these functions. By reconstituting the frequency spectrum of SNPs discovered in nine red foxes, we discovered an estimated 77–89% of all SNPs (within the region screened) present in North American red foxes. In sum, these findings indicate that information from the dog genome enables easy ascertainment of random and gene-linked SNPs throughout the Canidae and illustrate the value of SNPs in ecological and evolutionary genetics.

Keywords: *Canis latrans*, coyote, grey fox, red fox, single nucleotide polymorphisms, SNPs, *Urocyon cinereoargenteus*, *Vulpes vulpes*

Received 14 January 2007; revision accepted 13 April 2007

Introduction

Many questions in population genetics and molecular ecology depend on characterizing genome-wide variation within and between populations. For example, accurate reconstruction of historical demography or phylogenetic patterns requires multiple selectively neutral markers with

independent genealogies (Knowles & Maddison 2002). Detecting the effects of natural selection also requires a suite of selectively neutral genes to bound null expectations (Lewontin & Krakauer 1973; Beaumont 1996; Luikart *et al.* 2003; Storz & Dubach 2004; Storz 2005). In the past, autosomal microsatellites have been used to estimate neutral genetic divergence. Microsatellites occur throughout the nuclear genome and are highly polymorphic, making them valuable tools for qualitative assessment of population genetic divergence. However, due to homoplasy, genetic

Correspondence: Benjamin N. Sacks, Fax: +1 530 752 3556; E-mail: bnsacks@ucdavis.edu

distance indexes based on these markers will tend to underestimate population divergence and do so proportionally more as divergence times increase (Garza *et al.* 1995; Nauta & Weissing 1996; Estroup *et al.* 2002).

Single nucleotide polymorphisms (SNP) are scattered throughout the nuclear genome, especially in noncoding regions, and in principle are ideal for assessing neutral genetic divergence in phylogenetics and population genetics studies (Brumfield *et al.* 2003; Morin *et al.* 2004). SNPs, like microsatellites, are biparentally inherited (those on autosomes). However, in contrast to microsatellites, the mutation rates of these markers are several orders of magnitude lower, and the mutational process of SNPs is much better understood and easier to model, both of which provide the basis for more robust inferences (Kuhner *et al.* 2000; Wakeley *et al.* 2001; Morin *et al.* 2004; Nielsen *et al.* 2004). The low mutation rate of SNPs also should make them well-suited for interspecific phylogenetic studies (Brumfield *et al.* 2003; Lindblad-Toh *et al.* 2005). In contrast to mitochondrial DNA (mtDNA), which has a faster mutation rate, the accumulation of SNPs between species is likely to be more nearly linearly related to divergence time, reducing the uncertainty associated with deeply rooted splits.

The primary downsides associated with SNPs appear to be the difficulty in discovery in nonmodel organisms and the ascertainment bias resulting from the discovery process (Morin *et al.* 2004). An increasingly popular approach to obtaining SNPs in nonmodel organisms is to use highly conserved exon sequences (i.e. from multiple model organisms) as priming regions from which to amplify less-well conserved (i.e. variable) intron sequences from a wide range of taxa (Lyons *et al.* 1997; Belfiore *et al.* 2003; Smith *et al.* 2005; Sprowles *et al.* 2006). For example, Aitken *et al.* (2004) used this targeted-gene approach to amplify intron sequences across a wide range of mammals and successfully amplified and sequenced 50% of attempted loci in canids, although they did not screen sequences for SNPs. A potential downside to this approach for some applications is in its nonrandom coverage of the genome (Zhang & Hewitt 2003). A second approach that should minimize this shortcoming is to use the most closely related model organism as a source of random sequence from which to choose candidate sequences (Adams *et al.* 2006). For example, Seddon *et al.* (2005) took advantage of the dog genome to find random SNPs for the grey wolf (*Canis lupus*) with a high success rate similar to that in dogs (Parker *et al.* 2004). Wolves and dogs, however, are phylogenetically indistinct (Vilá *et al.* 1997; Savolainen *et al.* 2002). It is therefore of interest to know the extent to which this and other model genomes can be used for similar purposes in more distantly related nonmodel species.

Here, we investigate the applicability of the dog genome to the discovery of SNPs in more distantly related canids: the grey fox (*Urocyon cinereoargenteus*), red fox (*Vulpes*

vulpes), and coyote (*Canis latrans*). The grey fox is the most basal canid, from which other foxes (including the red fox) and dog-like canids diverged > 10 million years ago (Ma) (Wayne *et al.* 1997; Lindblad-Toh *et al.* 2005). The red fox and the coyote probably diverged from 7 to 10 Ma, while the coyote and the dog (via the wolf) apparently diverged approximately 1 Ma. We investigated the efficacy of using the dog genome to discover SNPs in these canids and evaluated these markers with respect to locus heterogeneity, population-level differentiation, and comparison to microsatellites. We used interspecific and intraspecific comparisons, including previously published data for the dog and wolf, to assess SNP accumulation rate homogeneity over multiple timescales. We also assessed ascertainment bias stemming from a small discovery panel and used this information to reconstitute and compare the frequency spectra of SNPs in red foxes, dogs, and wolves.

Materials and methods

Laboratory procedures

Locus selection, amplification, and sequencing. Our purposes required that SNPs be representative of the entire genome, ideally, a random sample. In practice, truly random selection of SNPs may be inefficient. Therefore, we used three approaches varying in their degree of randomness: (i) targeted genes ($n = 10$) shown previously to amplify introns in a canid (Aitken *et al.* 2004) (clearly nonrandom), (ii) a subset of an approximately random sample of dog bacterial artificial chromosome (BAC) clones ($n = 40$) but known to contain SNPs in the domestic dog and wolf (Parker *et al.* 2004; Seddon *et al.* 2005), and (iii) a randomly selected sample of sequences chosen from BAC clones ($n = 10$) obtained by the Broad Institute (<http://www.broad.mit.edu/mammals/dog/snp/>).

Primers for the 60 loci are shown in Table S1 (Supplementary material). Those for the 10 Broad Institute loci were designed by us using PRIMER 3 software (Rozen & Skaletsky 1998). We used a random number generator in Microsoft Excel to sort the available BAC clones in random order and chose one clone at a time, from which to design primers to amplify ~500 bases of either end of the BAC clone. Once primers were selected, we attempted to amplify via polymerase chain reaction (PCR) and sequence each DNA segment in a red fox, a coyote, and a grey fox. We extracted DNA from muscle or ear tissue specimens using the DNeasy tissue kit (QIAGEN). A universal PCR recipe was used for all primer pairs in a 23- μ L reaction volume including 1 \times Abgene PCR buffer IV, 2.8 mM MgCl₂, 0.2 mM dNTPs, 1 \times bovine serum albumin, 0.2 μ M primers, and 1 U Abgene *Taq* polymerase. The thermocycler profile was the same as that used by Aitken *et al.* (2004), except that it began with an initial denaturation step at 94 °C for

10 min (template DNA only), followed by 10 min at 85 °C during which time the remaining PCR reagents were added before beginning cycles. Because we wished to eventually multiplex loci, we made no attempt to optimize PCR conditions for loci not amplifying under these initial conditions. PCR was conducted in 96-well plates. A subset of 12–24 products was run on a 1.2% agarose gel and visualized via post-staining with ethidium bromide under ultraviolet light. If most products exhibited a single band, the entire 96-well plate was sequenced.

Sequencing involved purification of PCR product using Millipore PCR purification plates and a sequencing reaction using the BigDye Terminator version 2.0 Cycle Sequencing Kit (Applied Biosystems). In all cases, sequences used the forward primer (Table S1). Sequencing product was cleaned up using Millipore SEQ96 plates and then electrophoresed on an ABI 3730 capillary sequencer (Applied Biosystems). Sequences were aligned visually using SEQUENCHER 4.5 software.

Resequencing in a red fox panel. We chose those loci that produced clean sequences to resequence in a panel of nine red foxes and screened these for variable sites (i.e. SNPs). The number of foxes, nine, was chosen to minimize costs associated with a larger panel while retaining reasonable power to detect SNPs. For example, nine foxes (i.e. 18 gene doses) are sufficient to ensure 98% probability of detecting a SNP with minor allele frequency 20%. The nine foxes were chosen to be representative of the species throughout North America and included three groups presumed to be phylogenetically distinct: native western-mountain, eastern North America, and Alaskan/Canadian (Aubry 1983; Sacks, unpublished data). The ancestry of these foxes was assessed using mitochondrial cytochrome *b* and *d*-loop sequences and comparison to > 300 specimens from throughout North America (Perrine *et al.* 2007; Sacks, unpublished data). In a few cases, we used up the supply of DNA for an individual before testing all loci, in which case we substituted DNA of another individual from the same population. We PCR-amplified, sequenced, lined up sequences, and surveyed for polymorphic sites as described above.

Genotyping assay. Because the technology for genotyping SNPs is rapidly progressing, we did not emphasize genotype assay development in this study. Nevertheless, we developed and tested single base extension (SBE) primers for a subset of SNPs using an approach similar to Seddon *et al.* (2005). We chose one SNP at random per amplicon (i.e. amplified fragment) from a subset ($n = 13$) of polymorphic sequences among the nine red foxes for which to design adjacent primers for SBE genotyping reactions. The SBE primers were designed (using PRIMER 3 software) to sit directly adjacent to and upstream (5') of the SNP. Briefly, the SBE reaction is a one-step sequencing

reaction that works by adding a single, uniquely fluorescently labelled dideoxynucleotide, which is complementary to the SNP, to the 5' end of the SBE primer. A heterozygote exhibits peaks corresponding to two different dyes, whereas a homozygote shows one peak corresponding to a single dye. The product is then electrophoresed and genotyped according to which fluorescent dyes are present. By using primers of varying length, multiple SNPs can be genotyped in the same reaction and separated via electrophoresis in terms of their length in base pairs (Belfiore *et al.* 2003). Although we made no significant attempt to multiplex markers in this study, we incorporated poly(A) tails of varying lengths to assess the feasibility of doing so in the future. The SBE primers were used to blindly genotype eight of the nine original foxes for which sufficient amounts of DNA were available, after which genotypes were tested against the known sequences. The SBE genotyping were conducted in reactions containing purified (as above) PCR product (i.e. the amplified segments) as the template, the SBE primers developed in this study, and uniquely fluorescently labelled dideoxynucleotides (ABI PRISM SNaPshot kit, Applied Biosystems), with published biochemical and thermal conditions (Belfiore *et al.* 2003). Excess reagents were removed via shrimp alkaline phosphatase before electrophoresis on an ABI 3730 capillary sequencer (Applied Biosystems). Electropherograms were visualized using STRAND version 2.2.30 (Veterinary Genetics Laboratory, University of California, Davis).

Microsatellite genotyping. We genotyped the red foxes at 14 microsatellite loci: AHT-133 (Holmes *et al.* 1995), CPH2 (Fredholm & Winterø 1995), C01-424, C04-140, C08-618, FH2001, FH2004, FH2010, FH2054, FH2088, FH2289, FH2328, FH2380, and FH2457 (Breen *et al.* 2001). Reactions were carried out in two multiplexes of seven loci each with primer concentrations ranging 0.15–0.80 μM and a 17- μL volume. Reactions included 1 \times Abgene PCR buffer IV, 2.5 mM MgCl_2 , 0.2 mM dNTPs, 1 \times bovine serum albumin, and 0.7 U Abgene *Taq* polymerase. Forward primers were fluorescently labelled (6-FAM, VIC, NED, PET; Applied Biosystems). Reactions were carried out with the following temperature profile: 95 °C for 10 min, 85 °C for 10 min, 33 cycles of 95 °C for 1 min, 62 °C for 30 s and 72 °C for 45 s, and a final 30-min extension at 72 °C. The PCR products were electrophoresed using an ABI 3730 capillary sequencer in conjunction with an internal size standard, GENESCAN 500 LIZ (Applied Biosystems). Alleles were scored using STRAND version 2.2.30 (Veterinary Genetics Laboratory, University of California, Davis).

Data analyses

Interlocus variability in SNP density. The nuclear genome is heterogeneous with respect to the rate at which selectively

neutral mutations accumulate (Brumfield *et al.* 2003; Zhang & Hewitt 2003; Lindblad-Toh *et al.* 2005). We examined variability in both intra- and interspecific SNP density among loci. Because intraspecific SNPs are expected to compose a small proportion of bases (e.g. Parker *et al.* 2004), we compared the observed distribution of per-locus SNP density to Poisson expectations using a chi-squared goodness-of-fit test (Zar 1999). Because interspecific SNPs are likely to be prevalent (e.g. Lindblad-Toh *et al.* 2005), we assessed the variance in interspecific SNP density against binomial expectations by comparing observed values to 95% confidence intervals calculated according to locus-length in base pairs (Zar 1999).

Assessing selective neutrality of SNPs and microsatellites. To assess the possibility that certain loci might be under selection or linked to genes under selection, as well as to compare the performance of SNPs to microsatellites in gauging neutral population genetic divergence, we used the approach of Beaumont & Nichols (1996). Estimates of F_{ST} across the three phylogenetic groupings of red foxes were calculated based on SNP data and microsatellite data using β (Cockerham & Weir 1993) and weighted by heterozygosity across loci (Weir & Cockerham 1984). We then performed 20 000 coalescent simulations conditioned on these estimates and sample sizes under assumptions of the island model to generate 95% confidence regions bounding the relationship between F_{ST} and heterozygosity, against which the empirically estimated locus-specific F_{ST} -heterozygosity values were plotted. We used the program `FDIST2` for all of the above procedures (Beaumont & Nichols 1996).

Ascertainment bias and comparison of frequency spectra across species. A consequence of using a small number of individuals to discover SNPs is that SNPs with rarer minor alleles will tend to go undiscovered. The extent of this bias therefore depends on the demographic history of the population (and hence the preponderance of rare alleles) and the number of individuals used in the discovery set. Once SNPs have been discovered, a larger number of individuals is often used to screen for particular SNP alleles identified during the discovery process. Although larger screening sets increase the precision of SNP allele-frequency estimates, they do not affect the discovery bias. This form of ascertainment bias can be quantified and corrected via the frequency spectrum, which is a frequency distribution of SNPs based on their population allele frequencies (either ancestral or minor; Nielsen *et al.* 2004). In addition to enabling estimation of the ascertainment bias, the frequency spectrum summarizes all demographic information contained in the SNP data and is therefore an important parameter in its own right (Nielsen *et al.* 2004).

We used two approaches to compare frequency spectra of red foxes to previously studied canids, dogs and grey

wolves: (i) direct comparison of similarly biased frequency spectra (i.e. those based on SNPs discovered in eight or nine individuals), and (ii) comparison of reconstituted or unbiased frequency spectra. The dog data set used a large SNP discovery panel (also the screening panel) of 120 individuals (Parker *et al.* 2004). This allowed direct estimation of the unbiased frequency spectrum, but required us to estimate the biased frequency spectrum for comparison (i.e. the reconstitution procedure in reverse). We used an approach similar to that of Nielsen *et al.* (2004) although we binned frequencies to accommodate small data sets and assumed a 'folded' frequency spectrum (i.e. we did not differentiate between ancestral and derived alleles). Specifically, we estimated the number of SNPs expected to be observed in each of five minor-allele-frequency bins, k , (0–0.100, 0.101–0.200, 0.201–0.300, 0.301–0.400, 0.400–0.500), as a function of the number of chromosomes, d , used in the discovery set for all SNPs, i , with $\hat{p} \in k$ as

$$\hat{N}_{obs_k} = \sum_i \hat{Pr}(Asc_i | p) \quad (\text{eqn 1})$$

and

$$\hat{Pr}(Asc_i | p_i = \hat{p}_i) = 1 - (1 - \hat{p}_i)^d - (\hat{p}_i)^d, \quad (\text{eqn 2})$$

where p was the proportion of chromosomes with SNP i (the minor allele) in the set of 120 dogs from Parker *et al.* (2004) and Asc denoted ascertainment of the SNP in a sample of d chromosomes, 18 in the present case (i.e. 18/2 = 9 individuals).

Next, we reconstituted the red fox and wolf frequency spectra (both derived from small discovery sets, $n = 8$ or 9 individuals) for comparison to the unbiased dog frequency spectrum. To reconstitute the frequency spectrum in wolves, we estimated p_i for each SNP as its proportion (the minor allele) in the screening set of 108 Scandinavian wolves (216 chromosomes) and calculated the expected number of total SNPs (i.e. ascertained or not) in each of the five minor-allele-frequency bins based on the number of chromosomes used in the discovery set (i.e. d), using the subset of SNPs, i , for which $\hat{p} \in k$ as

$$\hat{N}_{reconst_k} = \sum_i \frac{1}{\hat{Pr}(Asc_i | p_i)}. \quad (\text{eqn 3})$$

We were able to estimate precise frequencies of each of the SNPs in the wolf population from the large screening set, which enabled reconstitution of an unbiased frequency spectrum. However, because no large screening set was available for precise estimation of allele frequencies in the red fox, we could not directly estimate an unbiased frequency spectrum. Instead, we reconstituted the frequency spectrum in red foxes in two ways expected to overestimate and underestimate rare allele frequencies, and used these estimates to bound an uncertainty interval. Our low estimate

was calculated for red foxes similarly to that for wolves, except that we estimated p_i for each SNP as its proportion (the minor allele) in the discovery set of nine foxes (18 chromosomes). This approach should have tended to underestimate the number of missed SNPs because the observed prevalence of SNPs, p_i was constrained to be $\geq 1/18$, although true prevalence in the population could have been lower. Our high estimate assumed that the frequency spectrum in the population within the low range of minor allele frequencies ($\hat{p} < 0.10$) was proportionally the same as for the Parker *et al.* (2004) dog data (i.e. based on 240 chromosomes). This approach would tend to overestimate the number of missed SNPs if dogs contained a greater proportion than red foxes of recent (i.e. rare) mutations, as has been found with other markers, consistent with the significant demographic increases in dogs (e.g. Vilá *et al.* 1997; Savolainen *et al.* 2002).

Results

Sequencing success

We attempted to sequence 60 homologous DNA loci (i.e. fragments), of which 48 fragments averaging 301 bp in length were successfully sequenced in at least one of the three canid species (Table 1). We sequenced 43 segments (13 043 bases total) in the red fox, 41 segments (12 390 bases) in the grey fox, and 42 segments (12 641 bases) in the coyote. Of the targeted genes (loci 1–10), six of 10 were sequenced in each species, whereas 37 of the 40 BAC-end sequences used previously to find SNPs in dogs (loci 11–50) were sequenced in at least one species. Only five of the randomly selected BAC-end sequences (loci 51–60) were successfully sequenced, and only two yielded clean sequences in more than one species. Post-hoc *in-silico* PCR (<http://genome.cse.ucsc.edu>) indicated that in four cases, these loci (or homologous sequences) were repetitive in the dog genome (loci 52, 53, 55, 57), expected to produce multiple PCR products, which largely explained the poor sequencing success in these randomly chosen loci. Because each amplification was only attempted a single time (i.e. no attempt was made to optimize PCR conditions), our success rate (approximately 75%, overall) was likely an underestimate of what was possible. Sequences were deposited in Genbank (EI522236–EI522401).

SNP discovery within the red fox and SNP evaluation

Of 40 segments (12 063 bases) successfully screened for SNPs in the panel of nine red foxes (average 8.4 individuals per locus, range = 6–9), 25 yielded a total of 46 SNPs or one for every 262 bases sequenced (Table 2). The 46 SNP found in red foxes included two insertions/deletion, 32 transitions, and 12 transversions. The number of intraspecific

substitutions per locus varied primarily within the bounds expected by chance. Numbers of SNPs per locus did not deviate significantly from Poisson expectations ($\chi^2_3 = 3.47$, $P = 0.32$). Unless otherwise stated, subsequent analyses are based on substitutions only.

To further explore the properties of intraspecific SNPs with respect to their representation of the red fox genome (in North America), we compared divergence estimates among the three population groupings (i.e. native western-mountain, eastern North America, and Alaskan/Canadian) among loci and relative to locus heterozygosity (Fig. 1a). All but one estimate fell within the range of selectively

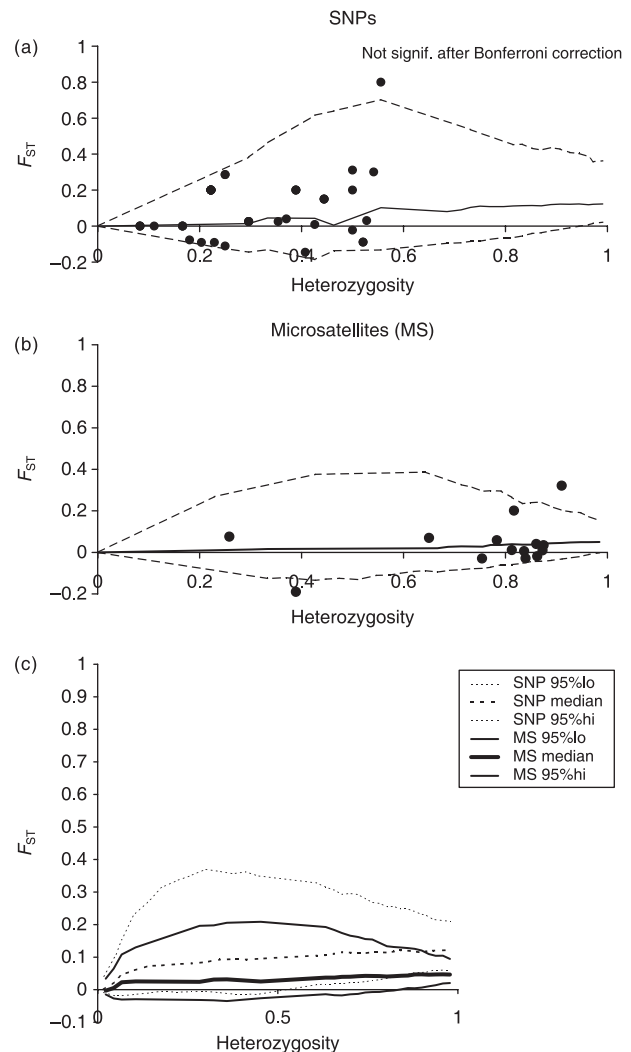


Fig. 1 Single-locus estimates of F_{ST} among three North American red fox populations relative to heterozygosity in reference to median expectations and 95% confidence bounds under the null hypothesis, no selection, for (a) single nucleotide polymorphisms (SNPs) and (b) microsatellites (MS); (c) comparison of SNP-based and microsatellite-based confidence regions, illustrating the bias associated with comparison of an SNP-based F_{ST} estimate to expectation based on microsatellites.

Table 1 Sequencing success for 60 independent segments of the genome in three canid species

Name	Amplicon ID	Length (bp)	Species‡	RC (no. of substitutions)	RG (no. of substitutions)	CG (no. of substitutions)	CD (no. of substitutions)
BGN	Locus 1	—	—	—	—	—	—
CFTR	Locus 2	473	RGC	13	10	12	—
CGA*	Locus 3	171	RGC	3	2	1	—
CK	Locus 4	—	—	—	—	—	—
COL9A1	Locus 5	256	RGC	4	4	4	—
FGFR4	Locus 6	351	RGC	2	4	4	—
GLUT2	Locus 7	334	RGC	2	2	2	—
GRP78	Locus 8	—	—	—	—	—	—
HOXD@	Locus 9	—	—	—	—	—	—
SCN4A	Locus 10	370†	RGC	8	8	6	—
372-C15-S	Locus 11	265	RGC	7	6	7	0
372-C15-T7	Locus 12	278	RGC	5	10	11*	0
372-C5-T7	Locus 13	301	RGC	4	5	7	1
372-E13-T7	Locus 14	372	RC	4	—	—	0
372-E2-SP6	Locus 15	265	RGC	5	5	2	1
372-I23-T7	Locus 16	170	RGC	3	2	3	0
372-M18-S	Locus 17	265	RGC	4	3	2	0
372-M23-T7	Locus 18	275	RGC	6	6	6	1
372-M7-T7	Locus 19	287	RG	—	2	—	—
372-M9-T7	Locus 20	291	RGC	9	10	9	1
373-A10-SP6	Locus 21	270†	RGC	6	14*	13*	0
373-A17-T7	Locus 22	279	RGC	5	8	5	0
373-A21-SP6	Locus 23	295	RGC	1	4	3	0
373-C13-SP6	Locus 24	461†	RGC	6	6	4	0
373-C15-T7	Locus 25	272	RGC	4	3	5	0
373-G19-T7	Locus 26	338	RGC	0	0	0	0
373-I16-SP6	Locus 27	—	—	—	—	—	—
373-I23-T7	Locus 28	268	RGC	4	8	8	0
373-I8-SP6	Locus 29	341	RGC	6	2	6	0
373-K8-SP6	Locus 30	304	RGC	4	5	3	0
372-C5-S	Locus 31	278	RGC	4	5	3	2
372-E15-S	Locus 32	208	RC	6	—	—	—
372-E15-T7	Locus 33	—	—	—	—	—	—
372-E18-T7	Locus 34	—	—	—	—	—	—
372-G17-T7	Locus 35	304	RGC	9	3	6	0
372-I23-SP6	Locus 36	447	RGC	14	11	5	0
372-M18-T7	Locus 37	330	RGC	1	4	3	0
372-M6-S	Locus 38	259	RGC	9	5	6	2
372-M6-T7	Locus 39	272	RG	—	6	—	—
372-O13-SP6	Locus 40	286	RGC	2	0	2	1
373-A14-T7	Locus 41	379	RGC	9	5	7	2
373-A15-T7	Locus 42	270	RC	2	—	—	0
373-A21-T7	Locus 43	263	RGC	5	5	0	1
373-E1-T7	Locus 44	275	RGC	4	3	5	0
373-E21-S	Locus 45	269	GC	—	—	7	0
373-E21-T7	Locus 46	281	G	—	—	—	—
373-G7-S	Locus 47	263	RC	6	—	—	0
373-G7-T7	Locus 48	289	RGC	3	3	4	0
373-I16-T7	Locus 49	282	RC	2	—	—	0
373-K10-T7	Locus 50	347†	RGC	12	6	10	—
BICF237J54685	Locus 51	323	RG	—	9	—	—
BICFPJ691874	Locus 52	—	—	—	—	—	—
BICF233J33962	Locus 53	—	—	—	—	—	—
BICFPJ1427627	Locus 54	293	R	—	—	—	—
BICFG630J156760	Locus 55	—	—	—	—	—	—
BICFPJ735430	Locus 56	—	—	—	—	—	—
BICFPJ1424092	Locus 57	329	G	—	—	—	—
BICF235J61841	Locus 58	—	—	—	—	—	—
BICFPJ325657	Locus 59	348	C	—	—	—	—
BICFPJ227583	Locus 60	156	GC	—	—	7*	—

*significantly more substitutions than other loci (pooled), $0.01 < P < 0.05$; †Locus 21 was 385 bases long, but was only screened for SNPs for the first 270 bases in interspecific and 226 bases in intraspecific comparisons due to the occurrence of large deletions in some individuals making it difficult to line up. Locus 10 appeared homologous in the grey fox until 370 bases, after which differences were too extensive to line up. Locus 24 and Locus 50 were not compared between coyotes and dogs due to shorter sequences in the dog data set; ‡Species from the present study are indicated: red fox (R), grey fox (G), coyote (C); dog (D) sequences available from Parker *et al.* (2004) (loci 11–50) are shown in comparison to the coyote.

Table 2 Single nucleotide polymorphisms (SNP) found in a screening set of up to nine red foxes relative to dogs and grey wolves

Amplicon ID	No. of foxes sequenced (d)	Length (bp)†	No. of SNPs in dog	SNPs present in wolves?	No. of red fox SNPs	Red fox SNP position† (bases)
Locus 3	9	209	—	—	0	—
Locus 5	8	250	—	—	0	—
Locus 6	9	367	—	—	1	132 (A/C)
Locus 7	8	349	—	—	0	—
Locus 10	8	584	—	—	1	538 (G/T)
Locus 11	9	291	2	No	3	127 (A/G), 202 (C/G), 239 (C/T)
Locus 12	9	308	1	No	4	37 (A/G), 95 (A/G), 116 (A/G), 198 (C/T)
Locus 13	9	304	1	No	0	—
Locus 14	9	374	1*	No	1	23 (A/G)*
Locus 15	9	273	4	Yes	0	—
Locus 16	9	209	1	Yes	0	—
Locus 17	9	270	1	Yes	1	17 (C/T)
Locus 18	9	278	5	Yes	1	139 (C/T)
Locus 19	9	302	2	Yes	0	—
Locus 20	9	303	2	Yes	5	66 (A/G), 169 (C/G), 171 (C/T), 190 (C/T), 208 (C/T)
Locus 21	7	226	1	No	2	55 (C/T), 208 (G/T)
Locus 22	9	193	1	Yes	1	78 (G/T)
Locus 23	9	299	1	Yes	1	229 (C/T)
Locus 24	9	292	1	No	0	—
Locus 25	9	290	2	Yes	1	71 (A/G)
Locus 28	9	273	0	No	0	—
Locus 29	9	365	2	No	1	131 (A/T)
Locus 30	7	309	1	Yes	0	—
Locus 31	9	282	2	Yes	0	—
Locus 32	9	207	4	Yes	0	—
Locus 35	9	309	1	Yes	4	76 (A/G), 219 (C/T), 262 (C/T), 268 (A/G)
Locus 36	6	448	1	No	0	—
Locus 37	6	338	1	Yes	0	—
Locus 38	8	272	5	Yes	1	229 (C/T)
Locus 39	8	277	2	Yes	1	146 (C/G)
Locus 40	8	285	1	Yes	1	78 (C/T)
Locus 41	6	369	3	No	2	77 (A/G), 122 (C/G)
Locus 42	8	276	1	No	2	58 (T-), 105 (C/T)
Locus 43	9	273	1	Yes	1	178 (C/G)
Locus 44	9	278	2	Yes	1	145 (G/T)
Locus 47	9	274	2	Yes	1	83 (A/T)
Locus 48	9	293	1	No	2	101 (C/T), 267 (C/G)
Locus 49	8	314	4	No	3	61 (C/T), 217 (A/G), 259 (A/G)
Locus 50	7	358	2	No	4	26 (A/G), 96–99 (DEL), 105 (C/T), 314 (A/G)
Locus 54	8	292	—	—	0	—

†Length and base-pair position of SNPs do not correspond exactly to those for interspecific comparisons, e.g. due to insertions/deletions.

*Locus 14 was the only intraspecific SNP site in dogs that was also found in the red fox, but the bases differed between the red fox and the dog (C/T, 372-E13T-57, frequency = 0.004; Parker *et al.* 2004).

neutral expectations (95% confidence region), and the one estimate falling outside this range was not statistically significant after Bonferroni correction for multiple comparisons. Moreover, estimates associated with two other SNPs on the same DNA segment (Locus 11) as the outlier (i.e. physically linked) were well within the range of neutral expectation. Divergence estimates based on microsatellite loci were considerably lower than SNP-based estimates. The overall F_{ST} estimates were 0.12 and 0.05 based on SNPs

and microsatellites, respectively. Additionally, the corresponding confidence regions for selective neutrality were substantially lower when based on microsatellites than on SNPs (Fig. 1b, c).

Reconstituted frequency spectra and ascertainment bias

The observed frequency of SNP alleles in the nine red foxes (i.e. 18 chromosomes) differed somewhat from those of

wolves and dogs, based on similar numbers of individuals (Fig. 2). In particular, minor allele frequencies exhibited modes between 10% and 20% in red foxes but between 30% and 40% in both wolves and dogs. Additionally, the dogs showed a mode between zero and 10%, indicating a relatively high frequency of rare alleles, consistent with higher population growth in dogs relative to the two wild canids. The reconstituted frequency spectra more clearly convey this difference between dogs and the two wild canid species. A positive consequence of the low frequency of rarer SNPs (even after adjusting for detection bias) in red foxes was that most SNPs in the population represented by the panel should have been detected by the nine individuals (usually) used in our panel. Based on the uncertainty interval, an estimate of between 77% and 89% of North American red fox SNPs was discovered. The 42 SNPs detected in our panel of nine red foxes in the 40 loci used in all three studies compared favourably to the number expected in the same-sized panel of dogs ($n = 50$ SNPs) and is higher than that discovered in the similar-sized ($n = 8$ wolves) panel of wolves ($n = 21$ SNPs). The latter comparison likely reflects differences in effective population sizes of the two wild canid populations.

Intraspecific SNPs in coyote and grey fox

Based on heterozygous sites, a coyote and grey fox had similar numbers of intraspecific SNPs: nine and 12, respectively (Table 3). These numbers are comparable to the red fox, for which the range of heterozygous sites per individual varied from six to 14 in individuals for which > 90% of loci were sequenced ($n = 5$).

While many loci containing SNPs in the dog and wolf also contained SNPs in the red fox, coyote, and grey fox, rarely were the SNP sites the same across species (except for the dog and wolf; Seddon *et al.* 2005). For example, of 62 dog SNPs in the 34 loci successfully sequenced in red foxes, only one occurred in red foxes (Table 2). However, the two bases in the red fox SNP and those in the dog differed, suggesting that these SNPs reflect convergence rather than a shared mutation. Coyotes are much more recently diverged from dogs and wolves and therefore might be expected to share some detectable proportion of SNPs. Of 60 dog SNPs on 33 loci successfully sequenced in the coyote, only one SNP (1.67%) also occurred in the coyote (Table 3). However, based on the number of heterozygous sites in individual red foxes relative to the total number of SNPs estimated for the species in North America, ~15% of SNPs are expected to be apparent in an individual. Therefore, assuming a similar relationship in coyotes implies an estimate of ~10% SNPs shared between coyotes and dogs. A more robust estimate can be made based on the number of observed coyote SNPs also found in the dog, one of nine, which produces a similar estimate, 11.1%.

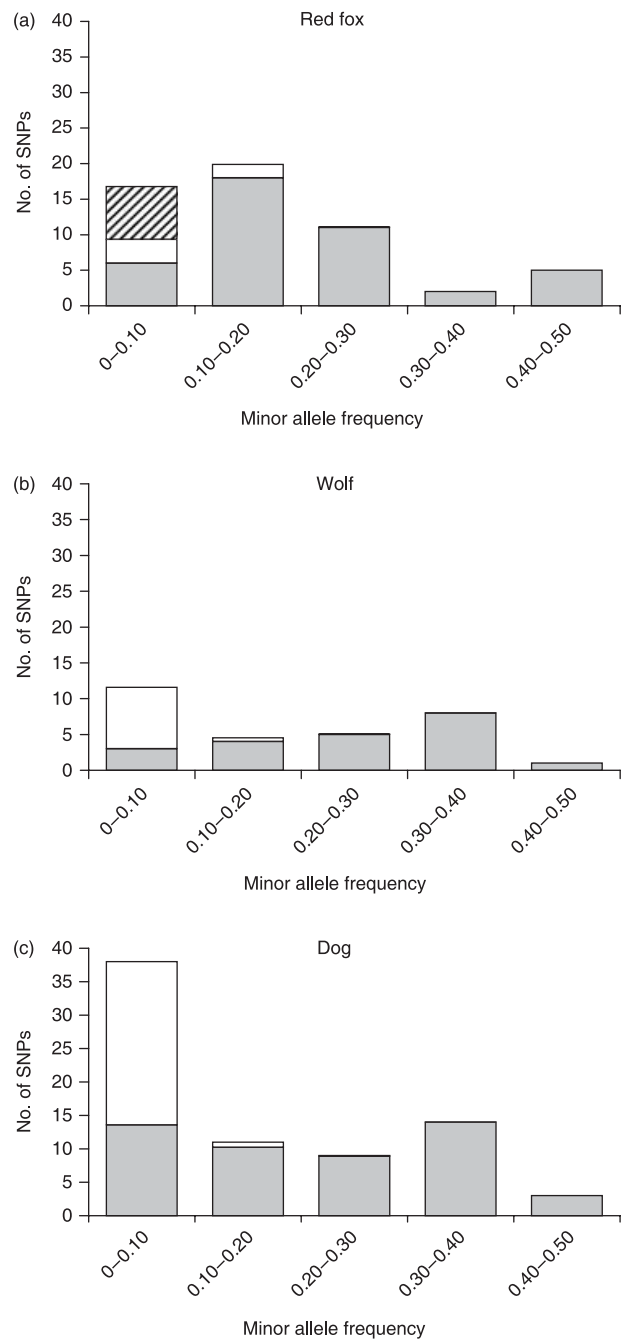


Fig. 2 Observed (shaded portion of bar) and reconstituted (total bar) frequency spectra for SNPs at 40 loci of three canid species as discovered in a panel of eight or nine individuals. An uncertainty interval for the reconstituted frequency spectrum of red fox (a) is indicated by cross-hatching (described in Materials and methods). Data for wolf (b) and dog (c) were from Seddon *et al.* (2005) and Parker *et al.* (2004), respectively. The 'observed' frequency spectrum for dogs corresponded to the expectation for a discovery panel of nine dogs as calculated analytically (detailed in Materials and methods) based on the allele frequencies observed in the much larger panel used by Parker *et al.* (2004). The 'reconstituted' frequency spectrum for dogs was the actual frequency spectrum observed in the full panel of 120 dogs used by Parker *et al.* (2004).

Table 3 Intraspecific single nucleotide polymorphisms (SNP) found in a single member of each of two canid species

Amplicon ID	Base no.	Coyote	Grey fox
Locus 2	277	—	G/T
Locus 2	418	—	A/C
Locus 2	419	—	A/G
Locus 2	461	—	A/C
Locus 2	462	—	A/C
Locus 6	172	A/G	—
Locus 17	191	—	A/G
Locus 17	228	—	A/G
Locus 19	149	—	C/T
Locus 21	236	A/G	—
Locus 22	220	C/G	—
Locus 24	165	—	C/T
Locus 24	358	—	G/T
Locus 26	322	A/G	—
Locus 36	64	A/C	—
Locus 36	209	A/G	—
Locus 44	84	—	A/G
Locus 44	251	—	C/G
Locus 48	203	A/G	—
Locus 49	94	C/G	—
Locus 49	253	C/T*	—

*The 253rd site at Locus 49 was also a C/T SNP in the dog (373-116T-302, frequency = 0.48; Parker *et al.* 2004). No other SNPs were shared with dogs or wolves (Seddon *et al.* 2005).

Interspecific SNP discovery and characteristics

All fragments for which at least two species were sequenced revealed one or more interspecific SNPs (Fig. S1, Supplementary material), by which we mean polymorphism was observed between, but not within, species. The substitution rate averaged 0.17% of bases per million years (Myr) (Table 4). Numbers of both transitions and transversions generally increased with species divergence time (Fig. 3). The phylogenetic relationships among the three distantly related species pairs cannot be resolved from substitution rates alone as apparent from the highly overlapping

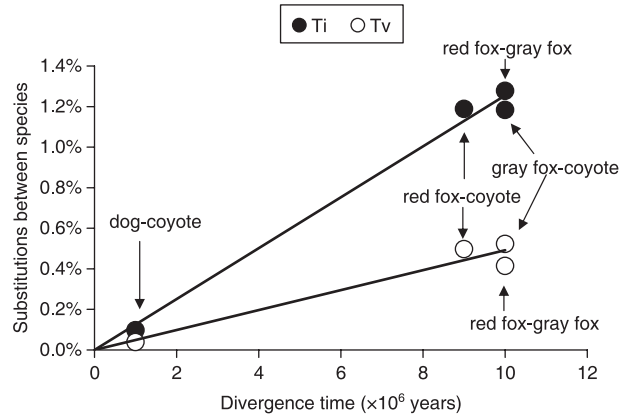


Fig. 3 Transitions (Ti) and transversions (Tv) vs. divergence time based on independent criteria (Wayne *et al.* 1997; Lindblad-Toh *et al.* 2005). Dog data from Parker *et al.* (2004).

confidence intervals (Table 4). However, a more sensitive indicator of the relationships, correlations in SNP density among fragments between species pairs (Fig. 4), does reflect the expected relationships based on independently derived knowledge of the phylogenetic relationships (Wayne *et al.* 1997; Lindblad-Toh *et al.* 2005). That is, the correlation in SNP density between red fox–grey fox and grey fox–coyote (Fig. 4a) was greater than that associated with either of the other two pair combinations (Fig. 4b, c). This is consistent with the more recent ancestry of the red fox and coyote relative to the grey fox, which implies that a portion of the mutations would have arisen in the red fox–coyote lineage before their divergence but after divergence with the grey fox, resulting in the observed higher correlation between the two species pairs that include the more ancestral canid.

In addition to shared ancestry, mutation rate heterogeneity among sets of homologous fragments originating in different parts of the genome (e.g. Lindblad-Toh *et al.* 2005) also would be expected to play into these correlations. On a locus-by-locus basis, the pairwise substitution rate between species varied from 0 to 0.33% per Myr (averaging 0.15% per Myr) but correspondence among species pairs was low

Table 4 Summary of polymorphisms found between sequences in pairs of canid species

Species pair	Indels	Substitutions	Total base pairs*	Substitution fraction (%) 95% confidence interval†
Red fox–coyote	20	182	10127	1.6–2.1
Red fox–grey fox	36	174	10127	1.5–2.0
Coyote–grey fox	26	173	10127	1.5–2.0
Coyote–dog	2	16	10741	0.1–0.2

*Only fragments sequenced in all three wild canids were included in these comparisons; coyote–dog comparisons are based on 34 fragments sequenced in both coyotes and dogs (dog sequences from Parker *et al.* 2004); †confidence intervals encompass the binomial probability density of numbers of substitutions (Zar 1999).

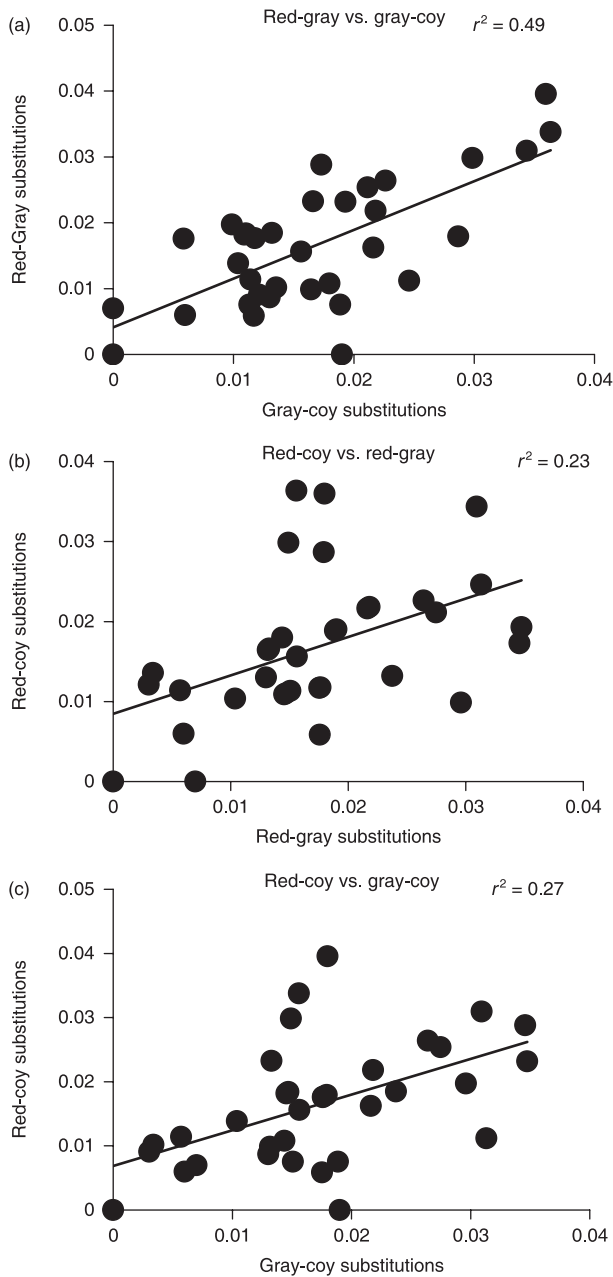


Fig. 4 Correlation between species pairs in SNP density (number of substitutions per base pair) among 44 homologous amplified nuclear DNA fragments: red fox–grey fox vs. grey fox–coyote (a), red fox–coyote vs. red fox–grey fox (b), and red fox–coyote vs. grey fox–coyote (c).

(despite the lack of independence associated with these pairwise data) and, as with the intraspecific SNPs, the variance among loci was low relative to sampling error expected within loci (Fig. 5). In fact, the variability across loci was sufficiently low as to be completely explainable by chance indicating, not that there was not heterogeneity in mutation rate, but — as with intraspecific SNPs — that this source of variance was essentially negligible.

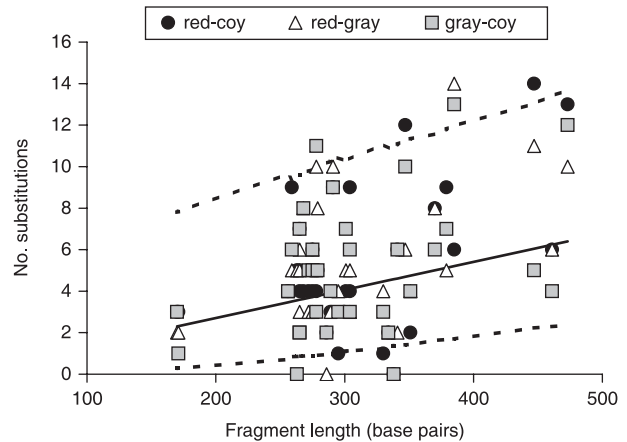


Fig. 5 Number of substitutions relative to fragment length observed in 34 loci amplicons sequenced in all three canids: red fox (red), grey fox (grey), and coyote (coy). For each species pair, four of 33 loci fell outside 95% confidence limits, which was expected to occur by chance with probability 0.06 (number of combinations of four from 33 times $0.05^4 \times 0.95^{29}$). Only one locus, 26, fell outside in all three species pairs; in all cases, there were 0 substitutions. Removal of this locus resulted in a probability of 0.15 that the remaining three loci fell outside confidence limits by chance.

Genotyping

SBE assays were designed for one SNP randomly chosen from 13 of the sequenced loci in red fox (Table 5). The SBE genotyping reaction produced readable genotypes in 83% of the cases, ranging from two to eight foxes per SNP locus. Of the readable genotypes ($n = 172$ gene doses), 97% agreed with the known genotypes based on sequencing. All five errors occurred in a single SNP locus (SNP 18–139).

Discussion

Theory and practice surrounding the use of SNPs to study population genetics of nonmodel organisms are rapidly developing areas of molecular ecology (Brumfield *et al.* 2003; Morin *et al.* 2004; Syvänen 2005; Dunbar 2006; Sanchez *et al.* 2006). More so than other markers, the process of ascertainment of SNPs can have important implications depending on the purposes of the markers. Targeted-gene approaches, which provide a means to amplify homologous sequences in known regions of the genome, are increasingly popular (Belfiore *et al.* 2003; Aitken *et al.* 2004; Smith *et al.* 2005; Sprowles *et al.* 2006). Linkage to genes of known function associated with this approach can be both an advantage and disadvantage, depending on the purpose. Such an approach is obviously ideal for studying particular genes. For uses such as parentage testing or population assignment, this type of approach also may be perfectly adequate (Morin *et al.* 2004;

Table 5 Single-base extension (SBE) primers designed for 13 loci and genotyping accuracy as measured against known sequences

Amplicon ID	SNP ID	SBE primer	No. of genes genotyped*	No. of genes correct
Locus 6	SNP6(132)-c.	A(×8)TCATCCTAGACCTCAGACATCT	16	16
Locus 10	SNP10(538)-gt	A(×12)ACTGTTACCACCTCTGC	12	12
Locus 11	SNP11(239)-ct	TGCTGTCTGGGAGCTCAG	16	16
Locus 12	SNP12 (198)-ct	A(×4)GGTGAACCTGGAAGCCATT	16	16
Locus 14	SNP14(23)-ct	A(×8)GGACACGGCTGCTGAGTA	16	16
Locus 17	SNP17(17)-ga	CTGGCAAATGAAAAAGCA	12	12
Locus 18	SNP18(139)-ga	A(×7)CATTTAAAGCACCTTTTGG	12	7
Locus 20	SNP20(66)-ga	A(×4)TACATGAGGTCTCCCAAT	16	16
Locus 21	SNP21(55)-ct	A(×18)AATGACATATTTTGCATGTAATA	14	14
Locus 22	SNP22(78)-gt	A(×16)AACTGTCACTAGCTGCACATA	14	14
Locus 23	SNP23(229)-ct	A(×14)AGGGCTGTCAGAAACCATTTC	14	14
Locus 25	SNP25(71)-ct	A(×22)TAACAAATTCATTTGTAGATTTAT	10	10
Locus 29	SNP29(131)-at	A(×26)GACTTCATTTTAGTGGTATCTGTG	4	4

*Assays were conducted for each locus on eight foxes but failed (i.e. produced no electrophoretic products in the expected range) for unknown reasons in 16 of 104 trials.

Seddon *et al.* 2005). However, for applications requiring unbiased representation of genomic variation, such as statistical phylogeography, population divergence, and to bound null expectations for the detection of genes under selection, randomly selected regions are more apt to have the desired statistical properties (Luikart *et al.* 2003). Finding randomly dispersed SNPs, however, presents additional challenges. One approach to random selection of genomic sequences in nonmodel organisms has utilized amplified fragment length polymorphisms [(AFLPs) Bensch *et al.* 2002; Nicod & Largiadèr 2003]. Ideally, randomly selected sequences homologous to those in model species can be used because of the associated information on linkage to functional genes that may prove important after locus selection. However, use of model genomes for this purpose requires that they be sufficiently similar to the target species for candidate priming sequences to exhibit high rates of cross-species amplification. To date, few studies have investigated random selection approaches based on model species (Primmer *et al.* 2002; Seddon *et al.* 2005; Adams *et al.* 2006). Consequently, it is difficult to assess the phylogenetic extent to which genome information from model species can be used to discover randomly dispersed SNPs in nonmodel relatives. Here, we investigated the usefulness on several levels of the dog genome as a tool for the development of SNPs for population genetic and phylogeographical studies in distantly related members of the Canidae.

Sequencing success

Our success at amplification and sequencing varied depending on how we selected primer pairs, with 60% vs. 80–88% of the targeted-gene loci and random loci from Parker *et al.* (2004), respectively, producing clean sequences

in the three nonmodel canids. The poorest success stemmed from the additional randomly selected BAC clones, due to amplification of multiple products repeated in the genome, as has occurred with birds (Primmer *et al.* 2002). In the future, however, candidate primers for randomly selected sequences could be checked against the dog genome via *in silico* PCR (<http://genome.cse.ucsc.edu>) before selecting loci (as we did here after the fact) to screen out repeatable elements. Amplification success was similar among nonmodel species despite the large range of times since divergence with dogs (Wayne *et al.* 1997). Overall, 68–72% of primer pairs produced sequences in the three canids examined here, which compares favourably to other studies. For example, only 57% of sequences derived from a DNA library of one bird species amplified in a different bird species within the same genus (Primmer *et al.* 2002) and 25–65% of gene-associated loci derived from targeted genes of multiple mammals amplified in 16 mammalian species (Aitken *et al.* 2004).

SNP discovery within species

The density of SNPs discovered in the red fox was nearly identical to what had been found in the wolf (Seddon *et al.* 2005) and within the range observed in a wide spectrum of taxa (Brumfield *et al.* 2003; Smith *et al.* 2005; Sprowles *et al.* 2006). Furthermore, the frequencies of heterozygous sites in a single grey fox and a coyote were comparable to those expected in a single red fox, wolf, or dog, suggesting a similar density in the other nonmodel canids. Given that the grey fox is the most ancestral of extant canids, this suggests that the dog genome can be used to find SNPs in any canid species with similar success.

On the other hand, no identical SNPs were found in dogs and red foxes, suggesting genotyping assays developed for

dogs are likely to perform poorly on red foxes and, by extension, on other distantly related canids. However, approximately 10% of SNPs appeared to be shared between dogs and the much less divergent coyote and, in another study, most SNPs in dogs were also found in wolves (Seddon *et al.* 2005). These estimates of shared SNPs are very crude but are generally in line with expectations according to relative divergence times. For example, assuming a generation time of 3 years in the wolf and 2 years in the coyote (Vilá *et al.* 1999), an average selectively neutral SNP would be expected to remain polymorphic for 1 Myr (the approximate divergence time between coyotes and dogs/wolves; Wayne *et al.* 1997), if the ancestral minor allele frequency were > 0.43 , for $N_e \sim 100\,000$ or > 0.12 , for $N_e = 500\,000$ in the wolf and $N_e = 150\,000$ or $750\,000$, respectively, in the coyote (calculated as per Kimura & Ohta 1969), all of which are plausible estimates (Vilá *et al.* 1999). For a neutral SNP to remain shared between red foxes and dogs, given the 10-Myr divergence time, would require effective population sizes of both species to be in the order of 1 million, which is unlikely, at least for modern-day wolves (Vilá *et al.* 1999).

Intraspecific SNP evaluation

Although the genome is heterogeneous with respect to mutation rate (Zhang & Hewitt 2003; Lindblad-Toh *et al.* 2005), intraspecific and interspecific SNP density variation among loci fell within the bounds expected by chance, i.e. due to sampling variance alone. Moreover, SNPs between species of known divergence times provided approximate estimates of transition and transversion rates, which were generally low, as expected for nuclear loci (Brumfield *et al.* 2003; Morin *et al.* 2004). These findings indicate that the magnitude and range of mutation accumulation rates in these markers should be negligible relative to drift and gene flow as effects on allele frequency divergence, which improves precision and statistical power associated with the use of these markers to represent genomic variation.

In contrast, microsatellites have relatively high mutation rates, which serves to obscure the effects of drift and gene flow on allele frequencies, and exhibit high levels of homoplasy and polymorphism, both of which are expected to bias estimates of population divergence (Garza *et al.* 1995; Beaumont & Nichols 1996; Nauta & Weissing 1996; Hedrick 1999). Indeed, we found that microsatellites systematically underestimated the null expectation of allele frequency differences relative to SNPs and can therefore be expected to produce an excess of type I errors if used to construct null distributions. Thus, it seems clear that relative to microsatellites, SNPs should provide more accurate estimates of population divergence due to drift and restricted gene flow, and provide a more accurate representation of the genome necessary for assessing deviations from neutral expectations of candidate loci thought to be under

selection. While true that a greater number of SNPs than microsatellites is required for most applications due to the bi-allelic form of the former marker (Mariette *et al.* 2002; Morin *et al.* 2004; Seddon *et al.* 2005), our findings demonstrate that, at least in the Canidae, a large number of SNPs can be easily found. Additionally, technological advances in genotyping methodologies are likely to make it increasingly feasible to assay large numbers of SNPs relatively cheaply (Morin *et al.* 2004; Schlötterer 2004; Syvänen 2005; Dunbar 2006; Sanchez *et al.* 2006).

Among SNPs and microsatellites, we found no statistical evidence that any of the loci were under selection or linked to genes under selection. An important caveat of these findings, however, is that our sample size (i.e. the number of foxes) was small and our statistical power to identify outlier loci was therefore low. It remains possible that a larger sample of foxes would reveal some loci to be statistical outliers consistent with selection or genetic hitchhiking and this will have to be assessed as part of future studies employing these markers.

Ascertainment bias and reconstituted frequency spectra

Our use of a discovery panel of nine red foxes to find SNPs necessarily led to an ascertainment bias with respect to the frequencies of SNPs. However, this bias appeared to be small due to the unimodal nature of the frequency spectrum. Because our panel spanned the major populations of North America, these SNPs should not be biased with respect to populations on this continent. On the other hand, North American red fox populations are derived from Eurasian populations such that the use of these SNPs for intercontinental comparisons can be expected to overestimate the diversity of North American populations relative to the more ancestral populations. For example, one study demonstrated that SNPs discovered from a derived population of *Drosophila melanogaster* incorrectly revealed similar diversity in the ancestral population but that SNPs discovered in the ancestral population more accurately revealed higher diversity in the ancestral than in the derived population (Schlötterer & Harr 2002).

Interspecific SNP discovery and characteristics

The relationship between interspecific SNP density and divergence time appeared approximately linear, for both transitions and transversions, suggesting these should be especially useful for phylogenetic studies (Brumfield *et al.* 2003). However, sampling variance dictates that SNP density alone will be a poor resolver of phylogenetic relationships at similar timescales, and it is therefore important to explicitly consider shared derived mutations as was apparent in correlations in SNP density per locus between species pairs in this study (see Fig. 4). Methods accounting for shared

derived mutations, primarily developed for mtDNA, can be used effectively with nuclear loci as well, and may be especially powerful if independent sets of linked SNPs are used (Hare 2001). Indeed, other studies employing nuclear SNPs have found greater phylogenetic resolving power in these markers relative to mtDNA sequences, consistent with the lack of homoplasy and replication associated with SNPs relative to mtDNA (Lindblad-Toh *et al.* 2005).

Prospects for genotyping assays

We made only preliminary efforts in this study to develop genotyping assays, mainly to assess how much effort would be required to do so. We designed SBE primers to work on a subset of our red fox SNPs, and in 12 out of 13 cases, these worked very well when used in single-plex reactions. However, migration relative to size standards in these short fragments varied considerably, which would have complicated interpretation of multiplex assays. In a single attempt to combine all 13 loci, we found that this problem, along with the occurrence of artifactual peaks, completely obscured our ability to interpret genotypes. Undoubtedly, we would have achieved greater success with smaller combinations of loci or by redesigning some SBE primers to minimize inhibitory primer interactions. For example, Seddon *et al.* (2005) was able to develop six SBE-based multiplex assays for 21 SNPs with purportedly little effort. Indeed, it may be possible to multiplex on the order of 50 loci in the SBE approach, as has been done with human SNPs (Sanchez *et al.* 2006). Alternatively, higher throughput methodologies enabling multiplexing of 50–150 loci on a single assay may be more time-efficient and cost-effective (Tobler *et al.* 2005; Dunbar 2006).

Conclusions

To date, few studies have explored the use of SNPs in nonmodel organisms. The present study extends the empirical support for the feasibility of such applications. In particular, our findings demonstrate that model species can be used effectively to develop SNPs for representing random genomic variation in nonmodel organisms > 10 Myr divergent, effectively the full phylogenetic range of canids. The high success rate we experienced in amplifying homologous regions among canids also implies that the dog genome should be an excellent source of gene-linked SNPs throughout the Canidae that can be compared to random SNPs to assess gene-specific selection. However, we observed a low transfer rate of SNP sites from dogs to other, especially more distant, canids, which suggests that particular SNP assays developed for dogs will have relatively low success rate as applied to the more distant canids, and species-specific assays will need to be developed in these cases. On the other hand, valuable information can

be gained by quantifying transferability of intraspecific SNPs from one species to another, which can be accessed via the use of gene chips developed for the dog that assay thousands of SNP sites (Lindblad-Toh *et al.* 2005; Syvänen 2005). For example, particular SNPs found to be shared among distantly related canids may mark genes associated with balancing selection. Thus, it seems clear that the dog genome holds tremendous potential for canid-wide applications of molecular ecology and evolution.

Acknowledgements

Funding was provided through a grant from the UC Davis Academic Federation Research Grant Program, Innovative Developmental Awards and through the Veterinary Genetics Laboratory (VGL) in the UC Davis School of Veterinary Medicine. We thank Niels Pedersen and VGL personnel too numerous to name for logistic and technical support, and are grateful to Tasha Belfiori, Amy Sprowles, Molly Stephens, Neil Clipperton, and Bernie May for helpful suggestions.

References

- Adams RI, Hallen HE, Pringle A (2006) Using the incomplete genome of the ectomycorrhizal fungus *Amanita bisporigera* to identify molecular polymorphisms in the related *Amanita phalloides*. *Molecular Ecology Notes*, **6**, 218–220.
- Aitken N, Smith S, Schwartz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology*, **13**, 1423–1431.
- Aubry KB (1983) *The Cascade red fox: Distribution, morphology, zoogeography and ecology*. Dissertation, University of Washington.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **263**, 1619–1626.
- Belfiore NM, Hoffman FG, Baker RJ, Dewoody JA (2003) The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Molecular Ecology*, **12**, 2011–2017.
- Bensch S, Åkesson S, Irwin DE (2002) The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Molecular Ecology*, **11**, 2359–2366.
- Breen M, Jouquand S, Renier C *et al.* (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Research*, **11**, 1784–1795.
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, **18**, 249–256.
- Cockerham CC, Weir BS (1993) Estimation of gene flow from F-statistics. *Evolution*, **47**, 855–863.
- Dunbar SA (2006) Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clinica Chimica Acta*, **363**, 71–82.
- Estroup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Fredholm M, Winterø AK (1995) Variation of short tandem repeats within and between species belonging to the Canidae family. *Mammalian Genome*, **6**, 11–18.

- Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution*, **12**, 594–603.
- Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.
- Holmes NG, Dickens HF, Parker HL *et al.* (1995) Eighteen canine microsatellites. *Animal Genetics*, **26**, 132–133.
- Kimura M, Ohta T (1969) The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, **61**, 763–771.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics*, **156**, 439–447.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Lindblad-Toh K, Wade CM, Mikkelsen TS *et al.* (2005) Genotype sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.
- Lyons LA, Laughlin TF, Copeland NG *et al.* (1997) Comparative anchor-tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, **15**, 47–56.
- Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within population diversity: trade-offs between markers. *Molecular Ecology*, **11**, 1145–1156.
- Morin PA, Luikart G, Wayne RK *et al.* (2004) SNPs in ecology, evolution, and conservation. *Trends in Ecology & Evolution*, **19**, 208–215.
- Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics*, **143**, 1021–1032.
- Nicod J-C, Largiadèr CR (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for nonmodel organisms. *Nucleic Acids Research*, **31**, e19.
- Nielsen R, Hubisz MJ, Clark AG (2004) Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*, **168**, 2373–2382.
- Parker HG, Kim LV, Sutter NB *et al.* (2004) Genetic structure of the purebred domestic dog. *Science*, **304**, 1160–1164.
- Perrine JD, Pollinger JP, Sacks BN, Barrett RH, Wayne RK (2007) Genetic evidence for the persistence of the critically endangered Sierra Nevada red fox in California. *Conservation Genetics* (in press).
- Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**, 603–612.
- Rozen S, Skaletsky HJ (1998) *PRIMER 3*. Code available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html.
- Sanchez JJ, Phillips C, Børsting C *et al.* (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, **27**, 1713–1724.
- Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science*, **298**, 1610–1613.
- Schlötterer C, Harr B (2002) Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. *Molecular Ecology*, **11**, 947–950.
- Seddon JM, Parker HG, Ostrander EA, Ellegren H (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology*, **14**, 503–511.
- Smith CT, Elfstrom CM, Seeb LW, Seeb JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology*, **14**, 4193–4203.
- Sprowles AE, Stephens MR, Clipperton NW, May BP (2006) Fishing for SNPs: a targeted locus approach for single nucleotide polymorphism discovery in rainbow trout. *Transactions of the American Fisheries Society*, **135**, 1698–1721.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.
- Storz JF, Dubach JM (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution*, **58**, 1342–1352.
- Syvänen AC (2005) Toward genome-wide SNP genotyping. *Nature Genetics*, **37** (Suppl.), S5–S10.
- Tobler AR, Short S, Anderson MR *et al.* (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *Journal of Biomolecular Techniques*, **16**, 396–404.
- Vilá C, Amorim IR, Leonard JA *et al.* (1999) Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus*. *Molecular Ecology*, **8**, 2089–2103.
- Vilá C, Savolainen P, Maldonado JE, Amorim IR *et al.* (1997) Multiple and ancient origins of the domestic dog. *Science*, **276**, 1687–1689.
- Wakeley J, Nielsen R, Liu-Cordero SN, Ardlie K (2001) The discovery of single nucleotide polymorphisms — and inferences about human demographic history. *American Journal of Human Genetics*, **69**, 1332–1347.
- Wayne RK, Geffen E, Girman DJ, Koepfli KP, Lau LM, Marshall CR (1997) Molecular systematics of the Canidae. *Systematic Biology*, **46**, 622–653.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Zar JH (1999) *Biostatistical Analysis*, 4th edn. Prentice Hall, Upper Saddle River, New Jersey.
- Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Ben Sacks uses molecular and field approaches to study behaviour, ecology, and evolution of terrestrial mammals. Susan Louie is a fourth-year undergraduate with a double-major in biochemistry/molecular biology and statistics, whose work on this project was part of an internship through the Canid Diversity and Conservation Project under the direction of Dr Sacks.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Single nucleotide polymorphisms found across 3 species in 44 independent homologous sequences. Ambiguity codes indicate heterozygous sites with bases corresponding to the Nomenclature Committee of the International Union of Biochemistry. Symbol ":" indicates a deletion; "-" indicates identity with the gray fox sequence; "not used" indicates an unsuccessful sequence. Dog sequence from Parker *et al.* (2004).

Table S1 Primers used to amplify sequences for SNP discovery in canids.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1471-8286.2007.01830.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.