Molecular Ecology (2008) **17**, 3808–3817 doi: 10.1111/j.1365-294X.2008.03876.x

To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations?

ÜLO VÄLI,*† ANNIKA EINARSSON,* LISETTE WAITS‡ and HANS ELLEGREN*
**Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36, Uppsala, Sweden, †Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia, ‡Department of Fish and Wildlife, College of Natural Resources, University of Idaho, Moscow, Idaho, USA*

Abstract

Microsatellite variability is widely used to infer levels of genetic diversity in natural populations. However, the ascertainment bias caused by typically selecting only the most polymorphic markers in the genome may lead to reduced sensitivity for judging genome-wide levels of genetic diversity. To test this potential limitation of microsatellite-based approaches, we assessed the degree of nucleotide diversity in noncoding regions of eight different carnivore populations, including inbred as well as outbred populations, by sequencing 10 introns (5.4–5.7 kb) in 20 individuals of each population (wolves, coyotes, wolverines and lynxes). Estimates of nucleotide diversity varied 30-fold (7.1×10^{-5}–2.1×10^{-3}), with densities of one single nucleotide polymorphism every 112–5446 bp. Microsatellite genotyping (10–27 markers) of the same animals revealed mean multilocus heterozygosities of 0.54–0.78, a 1.4-fold difference among populations. There was a positive yet not perfect (r^2 = 0.70) correlation between microsatellite marker heterozygosity and nucleotide diversity at the population level. For example, point estimates of nucleotide diversity varied in some cases with an order of magnitude despite very similar levels of microsatellite marker heterozygosity. Moreover, at the individual level, no significant correlation was found. Our results imply that variability at microsatellite marker sets typically used in population studies may not accurately reflect the underlying genomic diversity. This suggests that researchers should consider using resequencing-based approaches for assessing genetic diversity when accurate inference is critical, as in many conservation and management contexts.

Keywords: ascertainment bias, heterozygosity, microsatellite, nucleotide diversity, single nucleotide polymorphism

Received 10 March 2008; revision received 24 June 2008; accepted 25 June 2008

Introduction

Genetic diversity is a key factor enabling adaptation, and therefore survival, of natural populations in changing environments. Conversely, limited diversity may hamper the possibility for populations to adapt in the long term but loss of genetic diversity can also more immediately lead to decreased fitness within populations, due to inbreeding depression (Frankel & Soulé 1981; Lande 1988; Reed & Frankham 2003; Reed *et al*. 2003; Frankham 2005). Therefore, evaluations of genetic diversity are common in population genetics and are particularly important in conservation genetics (Frankham 2005). In principle, genetic diversity at loci with functional importance, such as protein-coding genes, RNA-coding or regulatory sequences, is what affects a population's ability to respond to selection. Ultimately, researchers would like to measure genetic variability across all or at least a significant fraction of such functionally important loci, an unrealistic scenario today but potentially possible within a few years.

While awaiting technology for large-scale genomic analysis of multiple individuals within a population, geneticists have primarily used neutral genetic markers to infer the levels and patterns of genetic diversity. If neutral markers are used, these should be expected to provide an

Correspondence: Hans Ellegren, Fax: + 46-18-4716310;
E-mail: Hans.Ellegren@ebc.uu.se

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

overall measure of the relative level of genomic variability (Hansson & Westerberg 2002). In the absence of balancing selection, e.g. selection for heterozygotes, neutral variability is typically higher than variability at functional loci subject to purifying selection. Although diversity levels vary among functional loci depending on the strength of selection, genetic diversity at neutral and functional loci should still be correlated since they are both dependent on effective population size (N_e) and other demographic parameters. Thus, neutral markers should be useful indicators of evolutionary relevant functional variability.

The first use of microsatellites in natural populations was reported more than 15 years ago (Ellegren 1991, 1992; Schlotterer *et al*. 1991). Since then, they have been the marker of choice for most applications in population genetics and molecular ecology. There is extensive knowledge on general aspects of their use (Selkoe & Toonen 2006), pitfalls (Pompanon *et al*. 2005), development (Zane *et al*. 2002), evolution (Ellegren 2004), and transfer across species (Barbara *et al*. 2007). Several properties make microsatellites attractive for studies of nonmodel organisms. From a practical point of view, they are relatively easy to develop and can be analyzed at moderate cost. From a conceptual point of view, the unusually high mutation rate of microsatellites renders them particularly useful in the context of estimating levels of genomic diversity. Since microsatellite mutation rates may be on the order of 10^{-3} per locus (Ellegren 2004), the parameter θ estimated from microsatellite data is typically orders of magnitude higher than corresponding estimates for unique DNA sequences, in which the point mutation per nucleotide may be 10^{-8}–10^{-9}. Because of this, significant amount of sequence data (in bp) may be needed to estimate mean nucleotide diversity with some confidence. In particular, when population diversity levels are low, finding a sufficient number of segregating sites in unique DNA sequences to estimate nucleotide diversity may require extensive sequencing. Clearly, this is less of a critical issue with microsatellites since at least some variation is usually seen with the marker sets typically employed.

However, the advantages provided by the high microsatellite mutation rate may produce a biased picture of the overall levels of genetic diversity. Variability at microsatellite loci may be seen as the genomic 'inverted tip of the iceberg', and we reason that this may not necessarily reflect genomic diversity, for several reasons. Strategies for microsatellite marker development almost always select for polymorphic or highly polymorphic loci, for example, because the longest repeat sequences within the genome are targeted during library enrichment and/or screening procedures. There is a well-known positive correlation between repeat length and microsatellite variability (Weber 1990; Ellegren 2000). Marker isolation is thus nonrandom with respect to diversity; regardless of the overall level of genomic variability, the most polymorphic loci are sought. Another concern is a potential publication bias in the application of microsatellite markers. Monomorphic markers or those with only limited variability in initial screenings may not be considered worthwhile to use in more extensive population surveys or in reports.

Surprisingly, few studies have addressed the relationship between population levels of genetic variability at microsatellite marker loci and at random DNA sequences. We have therefore designed a study to compare genetic diversity at these two types of sequence categories across several carnivore populations (of grey wolves *Canis lupus*, coyotes *Canis latrans*, Eurasian lynx *Lynx lynx*, and wolverine *Gulo gulo*), including small inbred populations as well as large outbred populations. We use data from microsatellite marker sets previously developed for these species and obtain estimates of neutral genomic diversity by multilocus resequencing of noncoding DNA regions.

Materials and methods

Samples

We analysed tissue samples of carnivores from four species, with a total of eight different populations worldwide: wolves from northwestern Europe (Sweden), eastern Europe (Tver and Smolensk region, Russia), southwestern Europe (Spain), North America (Great Lakes, USA), coyotes from North America (Great Lakes, USA), wolverines from North America (Montana, USA) and northwestern Europe (Scandinavia), and lynxes from northwestern Europe (Sweden). See references in Table S2, Supplementary material, for previous genetic studies of these populations, based on the same samples. From each population, 20 individuals were analysed.

Markers and laboratory methods

In each population, we studied sequence variability in 10 genes, usually in one intron from each gene (Table 1). We designed primers in exons flanking introns of an average of 550 bp in length, using either species-specific sequence data if available (i.e. the dog genome sequence for wolves and some wolverine gene sequences available in GenBank) or sequence information from closely related species (dog sequences for coyotes, mustelid sequences for wolverines; for lynx, primers developed for the other species were used). The 10 introns were in all cases from an equal number of genes that, according to information from the dog genome sequence (dogs were domesticated from wolves), are from different chromosomes and thus unlinked. The introns used partially overlapped across species; while some could be amplified in all four species, others were specific to one or two species. The introns were chosen without any prior information on their variability in the studied species. Two

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Table 1 A summary of the genes used for assessing nucleotide diversity

Symbol	Gene	Intron	Locus database ID
Aamp-2	Angio-associated, migratory cell protein	3	AY197355
Act	Alpha-actinin 4	5	ENSCAFG00000005776
Camk	Calcium/calmodulin-dependent protein kinase II gamma	1, 3	AY099467
C-fes	Feline sarcoma oncogene	12	AF498174
Chy	Chymase precursor	3	U89607
Des	Desmin	5	BK005142
Ghr	Growth hormone receptor precursor	8	AF498195
Hmgb1	High-mobility group B1 protein	3	AY135520S2
Nramp-1	Natural-resistance-associated macrophage protein	4	AF091049
Ntrk3	Neurotrophic tyrosine kinase receptor type 3	4	ENSCAFG00000011457
Nup	Nucleoporin Nup107	23	ENSCAFG00000000417
Pde6b	Phosphodiesterase 6B	17, 18	AJ278002
Rab	Rab GTPase-binding effector protein 1	15	ENSCAFG00000015589
Shc1	SH2 domain protein C1	7	ENSCAFG00000017145
Spink5	Serine protease inhibitor, Kazal type 5	14	AJ972677
Tyr	Tyrosine aminotransferase	6, 9, 11	AF163863

introns from the same gene were targeted in one case for lynxes and in two cases for wolverines.

Intron amplification by polymerase chain reaction (PCR) was performed in 25 L solutions containing 20–50 ng DNA, 0.75 U Ampli*Taq* Gold polymerase with Ampli*Taq* Gold PCR buffer (Applied Biosystems), 1.5–2.5 mM $MgCl_2$, 0.2 M of each primer and 0.2 mM dNTP. The PCR profile included initial heating at 95 C for 5 min, followed by 35 cycles of 94 C for 30 s, annealing temperature for 30 s and 72 C for 45 s, and a final extension at 72 C for 10 min. Locus-specific PCR details, including primer sequences are presented in Table S1, Supplementary material. PCRs were purified using ExoSAP-IT (Amersham Biosciences) and used for direct sequencing with a DYEnamic sequencing kit and run on a MegaBACE (Amersham Biosciences) instrument. Both strands were sequenced from each end in order to check potential errors and avoid ambiguities in sequences. Intron sequences are deposited in GenBank under Accession nos EU871592–EU871621.

Microsatellite data of the same individuals was obtained from previous or ongoing studies (Hellborg *et al*. 2002; Flagstad *et al*. 2003, 2004; Cegelski *et al*. 2006; A.-K. Sundqvist *et al*. submitted; M. Nord and J. A. Leonard, in preparation; see Table S2 for markers used). The number of markers used per population varied between 10 and 27.

Data analysis

Sequences were aligned and checked using Sequencher 4.6 (Gene Codes) and BioEdit 7.0 (Hall 1999). Nucleotide diversity was calculated by DnaSP 4.5 (Rozas *et al*. 2003) on the basis of allele sequences, i.e. counting two times each homozygous nucleotide call (making 40 alleles per population). We used the total number of single nucleotide polymorphisms (SNPs) and the total sequence length in our diversity calculations, but similar results were obtained when we used the average values of nucleotide diversities of 10 introns. Sequence diversity was also characterized by Waterson's θ estimator, defined as $\theta = S/(1 + 1/2 + 1/3 + ... + 1/n - 1)$, where S is the number of segregating sites and n is the sample size (Waterson 1975; Nei 1987).

Intron haplotypes were reconstructed for each locus separately, using all detected SNPs, by the program PHASE 2.1 (Stephens *et al*. 2001; Stephens & Scheet 2005). Comparatively, we mimicked an ordinary genotyping of loci by including only one (the most polymorphic) SNP per locus in heterozygosity calculations. Expected (in population-level analyses) and observed (individual-level analyses) heterozygosities, as well as mean number of alleles per locus, were obtained by Microsatellite Toolkit for Microsoft Excel (Park 2001). As distributions of genetic diversity variables are hard to predict, between-population comparisons were conducted using nonparametric Kruskal–Wallis tests. We used a simple linear regression and untransformed variables in order to describe the relationship between nucleotide diversity and microsatellite heterozygosity since transformation of variables did not straighten up the line and the number of studied populations was too small to support a more complicated relationship.

Results

We retrieved 5.4–5.7 kb of sequence data from 10 introns in 20 individuals each of four wolf populations, two wolverine populations, one coyote population and one lynx population. There was significant variation in levels of polymorphism among populations. The North American wolf and coyote populations each showed about 50 segregating sites or one

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Table 2 Polymorphism information for 10 introns sequenced in 20 individuals from each carnivore population. *S* is the number of segregating sites (number of singletons given in parentheses)

Population	Length (bp)	*S*	No. of variable introns	Waterson's θ (per sequence)	SNP density	
					SNPs per 10^3 bp	bp per SNP
Wolves, Sweden	5697	25 (3)	7	5.9	4.4	228
Wolves, Spain	5697	27 (4)	8	9.4	4.7	211
Wolves, Russia	5697	34 (2)	9	7.8	6.0	168
Wolves, Great Lakes	5697	46 (9)	10	10.8	8.1	129
Wolverines, Scandinavia	5446	3 (0)	1	0.7	0.6	1815
Wolverines, Montana	5446	1 (0)	1	0.2	0.2	5446
Coyotes	5697	51 (9)	9	12.0	9.0	112
Lynxes	5720	2 (0)	1	0.5	0.3	2860

Table 3 Estimates of nucleotide diversity (×10^3) for each gene and population

	Wolves				Wolverines		Coyote	Lynx
Gene	Sweden	Spain	Russia	USA	USA	Scandinavia	USA	Sweden
Aamp-2	0	0.37	0	0.96	0.64	0	0	0
Act	2.60	2.65	2.14	0.90			0.27	
Camk intron 1					0	0		
Camk intron 3	0.11	0.70	1.15	1.37			1.29	
C-fes	0	0	0.18	0.65	0	5.16	0.18	
Chy					0	0		
Des								0
Ghp								0
Hmg					0	0		
Nr	1.16	1.30	1.37	3.40	0	0	2.66	0
Nt	1.32	2.01	3.96	4.31			2.92	
Nup	2.35	6.15	4.78	5.45			6.67	0
Pde intron 17	0	0	0.61	2.33	0	0	3.41	
Pde intron 18					0	0		0
Rab	1.73	1.73	1.57	0.82			1.77	
Shc	1.58	0.57	2.20	0.90			1.01	1.03
Spink								0
Tyr intron 6								0
Tyr intron 9					0	0		
Tyr intron 11					0	0		0
Average	1.09	1.53	1.78	2.11	0.07	0.57	2.02	0.10

SNP every 112–124 bp in the 40 chromosomes sampled per population (Table 2). The wolverine and lynx populations were much less variable, with one to three polymorphic sites or one SNP every 1815–5446 bp. The European wolf populations showed intermediate levels of polymorphism, with about 30 segregating sites or one SNP per 168–228 bp. Nucleotide diversity (π), essentially the pairwise heterozygosity per bp, varied from 7.1×10^{-5} in North American wolverines to 2.1×10^{-3} in North American wolves, a 30-fold difference in diversity level (Table 3). The same populations also showed the lowest (0.03) and highest (0.53) values of SNP heterozygosity when calculated using reconstructed SNP haplotypes. Such haplotype-based heterozygosity was very strongly correlated with nucleotide diversity ($r^2 = 0.98$, $P < 0.001$). However, the correlation between nucleotide diversity and heterozygosity calculated using only the most polymorphic SNP per locus was weaker ($r^2 = 0.62$, $P < 0.05$).

Data from 10–27 microsatellite markers genotyped in the same individuals of all populations were obtained from previous or ongoing studies. Mean expected heterozygosity (H_E) for these markers ranged from 0.54 to 0.78, i.e. about 1.4 times higher in the most polymorphic population (coyotes) compared to the least variable (lynx). While

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

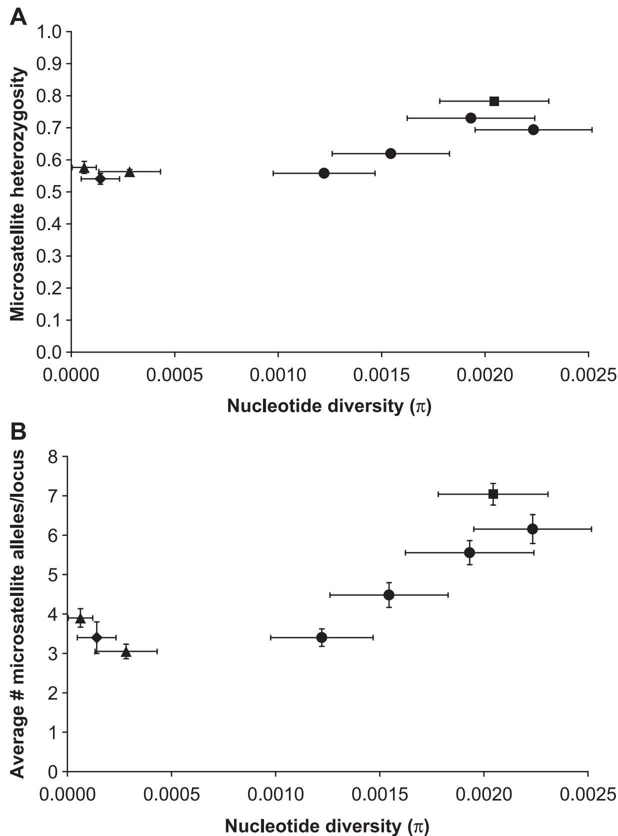

Fig. 1 Relationships between nucleotide diversity and (A) microsatellite heterozygosity and (B) mean number of microsatellite alleles in eight carnivore populations. Wolverine populations are indicated by triangles, wolf by circles, lynx by a diamond and coyote by a square. Whiskers indicate standard errors.

nucleotide diversity is an estimator of the parameter θ (θ = $4N_e$; where N_e is the effective population size and the mutation rate), H_E is not. However, assuming an infinite allele mutation model, $H_E/(1 - H_E)$ is a direct estimator of θ and can thus be used in comparisons to π. $H_E/(1 - H_E)$ ranged from 1.17 to 3.61, representing a threefold range of variation.

There was a significant positive correlation between nucleotide diversity and microsatellite heterozygosity among populations ($r^2 = 0.70$; $P < 0.01$; Fig. 1A), although not as strong as the correlation between nucleotide diversity and SNP haplotype heterozygosity presented above. Notably, the observed point estimates of mean microsatellite heterozygosity are essentially indistinguishable among the five least variable populations despite the fact that nucleotide diversity varies with an order of magnitude among these populations. A similar relationship was seen between nucleotide diversity and the mean number of microsatellite alleles per locus ($r^2 = 0.71$; $P < 0.01$) (Fig. 1B).

We also analyzed the relationship between SNP and microsatellite heterozygosities (proportion of heterozygous positions or loci) among individuals (Fig. 2); this analysis was restricted to the wolf and coyote populations where the number of markers was high and variability more extensive. Both microsatellite (Kruskal–Wallis test: $H = 32.3$, $P < 0.001$) and SNP-based ($H = 21.9$ for all 75 SNPs, $H = 31.7$ for reconstructed haplotypes, $P < 0.001$ in both tests) observed heterozygosities differed significantly among populations. However, there was no significant correlation between microsatellite and SNP heterozygosities in any of the five populations investigated, neither when all SNPs, reconstructed intron haplotypes or only the most polymorphic SNP per locus were used. The only trend of a positive correlation between microsatellite and SNP heterozygosities was seen in the Scandinavian wolf population ($r^2 = 0.09$, $P = 0.22$).

Discussion

This study has two important results. First, we confirm that, at the population level and as expected, there is a positive correlation between genetic diversity estimated by sets of microsatellite markers and by resequencing of multiple noncoding regions in the genome. Second, we show that the magnitude of variation in population genetic diversity is considerably higher for estimates of nucleotide diversity than for microsatellite heterozygosity. As a consequence, two populations with similar microsatellite heterozygosity may differ significantly in the overall levels of genomic diversity. At the individual level, we were unable to demonstrate a positive correlation between individual microsatellite heterozygosity and nucleotide diversity. Below, we discuss the consequences of these observations and, more generally, the factors that may affect estimates of genetic diversity in different types of DNA sequences.

Related studies

It is difficult to place our research in the context of the broader literature because there are few similar studies for comparison. Payseur & Cutter (2006) used coalescence simulations to model the effects of mutation and genealogical history on the correlation between microsatellite and SNP variability. They concluded that the correlation may often be weak due to, for example, large sampling variance and the complex nature of the microsatellite mutation process. Moreover, their simulations assumed no recombination between loci. In an empirical study, Ryynänen *et al*. (2007) recently studied the relationship between microsatellite heterozygosity and variability of nine SNPs/indels in Atlantic salmon. These parameters were correlated ($R^2 = 0.42$), and the range of individual heterozygosities was more similar between microsatellites and SNPs than in our study. Clearly, however, with the

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

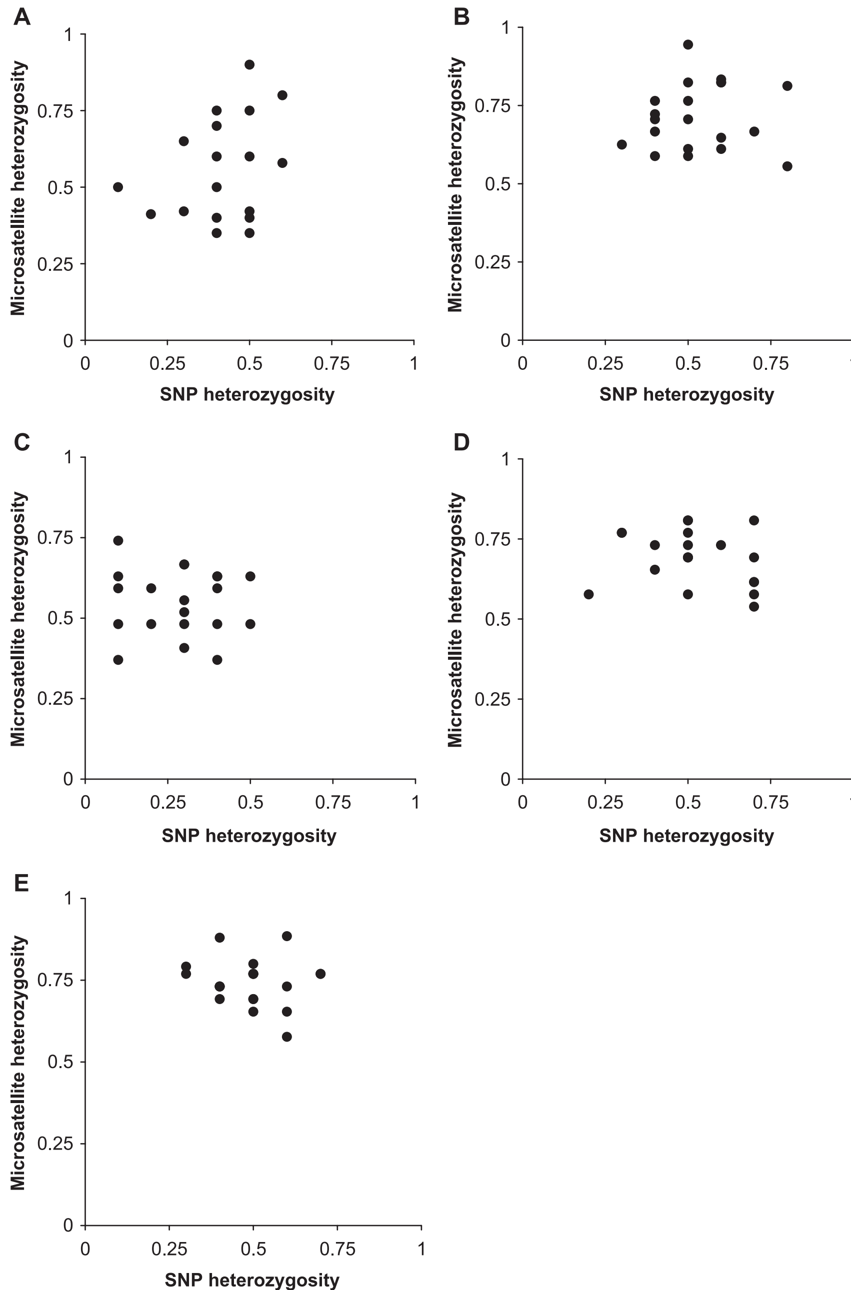

Fig. 2 Relationships between observed individual microsatellite and reconstructed SNP haplotype heterozygosities in (A) Swedish, (B) Russian, (C) Spanish and (D) Great Lakes wolf populations, and (E) Great Lakes coyote population.

focus on ascertained SNPs, the design of the study by Ryynänen *et al*. (2007) differs from our resequencing approach.

Microsatellite ascertainment bias

Microsatellite ascertainment bias usually means that markers tend to be more polymorphic and variable in the population from which they have been isolated (Ellegren *et al*. 1995, 1997). The concept of a microsatellite ascertainment bias is applicable also within genomes as marker isolation selects for the most polymorphic markers. Whole-genome analyses have shown that microsatellite markers used in population studies differ from random sets of microsatellite loci in several respects (Pardi *et al*. 2005; Brandström & Ellegren 2008), including, for example, length and incidence of repeat interruptions.

The focus on polymorphic markers in microsatellite isolation and screening inevitably biases estimates of the overall levels of genomic diversity. Mean multilocus microsatellite heterozygosities are typically larger than 0.30–0.40, and lower than 0.85, i.e. about a twofold difference between populations with low and high variability, respectively [corresponding to a ≈10-fold difference if

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

considering the estimator of θ, $H_E/(1 - H_E)$]. As shown here, the underlying difference in neutral DNA sequence heterozygosity between populations with even moderate variation in microsatellite heterozygosity (our range was 0.54 to 0.78) can be significant and in this case 30-fold. However, we note that a more quantitative analysis of the differences between θ estimated by microsatellites and resequencing, respectively, have to use more realistic mutation models for microsatellites, as $H_E/(1 - H_E)$ applies only with an infinite allele size model.

In principle, selecting polymorphic markers to measure diversity introduces an ascertainment bias regardless of which type of DNA sequence is analysed, and has been recognized as a confounding factor in studies based on SNPs as well (Nielsen 2004; Clark *et al*. 2005). It is therefore important to distinguish between analyses of variation in unique DNA sequence measured by SNP genotyping and by resequencing. The former approach focuses only on sites which are known to be polymorphic while the latter implies analyses of diversity in DNA sequences without prior knowledge of polymorphism. With the advent of new sequencing technology (Hudson 2008), we envision that sequencing surveys will gradually come to replace microsatellites in at least some types of genetic studies of natural populations. As a consequence, we foresee investigators will become increasingly familiar with thinking of population variability in terms of nucleotide diversity rather than in the form of mean microsatellite heterozygosity. This is not to say that microsatellites will no longer be useful for analyses of, for example, dispersal and population structure based on large sample sizes. However, sequencing-based approaches can provide more accurate and precise information on genomic levels of diversity, which is critical in assessing, for example, the genetic status of endangered populations (see further below).

Using neutral loci

Together with genetic drift, selection and mutation govern the amount of variation in a population. If the aim is to obtain an unbiased estimate of genomic diversity, it seems reasonable to focus on neutral sequence (in theory, variability at loci under selection may also be used to infer genomic diversity; however, for this to be meaningful and possible to compare among populations, the unrealistic assumption of constant selection over time and among lineages would have to be made). Microsatellites are often considered neutral markers. although there have been attempts to invoke general functional properties (Kashi *et al*. 1997; Li *et al*. 2004). However, there is an increasing number of reports on associations of allelic length variants at microsatellite loci and expression pattern of linked genes (reviewed in Kashi & King 2006); thus, selection cannot be completely excluded.

For noncoding sequence, like introns or intergenic DNA, the often-made assumption of selective neutrality is challenged by accumulating data on the important role played by regulatory sequences in phenotypic evolution (Carroll 2005). In fact, analyses of the human genome indicate that the majority of the genome that evolves by purifying selection is not protein-coding (for details, see the ENCODE Project Consortium 2007). Moreover, when N_e is large, as in *Drosophila*, positive selection can be demonstrated in many noncoding regions (Andolfatto 2005). The function of this genomic 'dark matter' remains largely unknown; however, recent data suggest that a significant part of the genome is actually transcribed into RNA (Pheasant & Mattick 2007). Whether this represents 'transcriptional noise' or perhaps indicates a biological function also remains unclear. With the exception of a tendency for the first intron of protein-coding genes and sites immediately flanking intron–exon junctions to be conserved (Chamary & Hurst 2004), it seems reasonable to assume that a random choice of intergenic and/or intronic DNA should provide an unbiased estimate of overall levels of genomic variability.

Even with the analysis of neutral or nearly neutral DNA sequences, selection may still affect diversity levels due to its action on linked loci. Background selection and, in particular, selective sweeps leave genomic footprints in the form of reduced diversity and extended haplotype structure at linked sites. The range of such effect is dependent on the rate of recombination, and thus, independent of type of DNA sequence. However, the rate of mutation will determine how fast the footprint of selection fades away. Thus, selection will have less long-lasting effects on microsatellite variability than on unique sequence (Wiehe 1998). Moreover, on the whole, since microsatellites often occur in intergenic regions, they may be less affected by selection compared to introns within coding sequence.

Within-genome and between-genome variation in mutation

There is extensive variation in point mutation rates within genomes (Ellegren *et al*. 2003), at scales from local sequence context effects (like the hypermutability of cytosine in methylated CpG dinucleotide sites) up to differences between chromosomal classes (like the different mutation rates of sex chromosomes due to male-biased mutation). Moreover, base composition and recombination rate are often correlated with the rate of point mutation. Because of this, for any given amount of sequence data collected for measuring genomic diversity, it is advisable to study several different regions rather than gathering data from equally many base pairs from a single locus (Brumfield *et al*. 2003).

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

For microsatellites, as many loci are usually analysed, any similarly systematic variation in mutation rates within genomes should in most cases be less of a concern, as long as markers are randomly distributed. However, there are circumstances when systematic variation in mutation rates can be important. Equilibrium microsatellite length distributions within genomes are thought to be governed by an intricate interplay between the rate of replication slippage and the rate of point mutation (Bell & Jurka 1997; Kruglyak *et al*. 1998). Due to the propensity for slippage mutations to lead to repeat expansions, coupled with the positive correlation between length and mutation rate, high slippage rate favours the evolution of long and polymorphic loci. In contrast, point mutations lead to repeat interruptions, known to decrease the rate of length mutation and hence polymorphism (Petes *et al*. 1997; Rolfsmeier & Lahue 2000; Sibly *et al*. 2003). An important consequence of this is that, within species, microsatellites located in genomic regions with a high point mutation rate should be less polymorphic, an assumption supported by empirical data (Santibanez-Koref *et al*. 2001; Brandström & Ellegren 2008). Excluding other variables, the overall level of microsatellite polymorphism may therefore be expected to be lower in species with a high rate of point mutation, as observed in *Drosophila melanogaster* (Schug *et al*. 1997, 1998). This may have been an overlooked problem in studies of natural populations as it means that, in theory, a population with high levels of genomic diversity may show lower levels of microsatellite variability than a population with less genomic diversity.

Another aspect of the relationship between microsatellite variability and genomic diversity is that mutability, and hence polymorphism, is influenced by the nucleotide composition of repeat motifs. Specifically, the replication slippage rate increases in most cases with decreasing GC content of repeat motifs (Brandström & Ellegren 2008), an intuitive consequence of the lower-strand stability of AT-rich sequence due to weak hydrogen bonds (Schlotterer & Tautz 1992). This means that the strategy for isolating markers, i.e. the use of particular motif probes in enrichment and/or hybridization, shall come to affect the polymorphism levels later observed.

Implications for conservation

The species under investigation in this study comprise carnivore populations that in many parts of the world have gone extinct or are endangered. Large mammalian predators have decreased due to habitat destruction and other types of human activity (Aspi *et al*. 2006; Linnell *et al*. 2007), and this has affected the genetic diversity and structure of the populations studied here (Vilà *et al*. 1999; Walker *et al*. 2001; Hellborg *et al*. 2002; Flagstad *et al*. 2003; Cegelski *et al*. 2006), and led to inbreeding depression (Liberg *et al*. 2005). It is therefore of vital importance to accurately assess genomic diversity and monitor how levels of diversity change. Under the assumption that there are no significant differences in mutation rates among species, our data indicate a very low historical N_e for North American wolverines and Swedish lynxes, and a low N_e for Scandinavian wolverines. This study thus clearly shows that microsatellite marker heterozygosity may miss or grossly underestimate significant differences in nucleotide diversity among carnivore populations. This should be taken as a cautionary observation and suggests that future studies of endangered populations should integrate DNA sequencing-based measures of genetic diversity.

Acknowledgements

We thank Neil Anderson, Jennifer Leonard, Maria Nord, Anna-Karin Sundqvist and Carles Vilà for providing samples and microsatellite data. Comments of Jennifer Leonard and three anonymous referees helped to improve the manuscript. Financial support was obtained from the Norwegian and Swedish Environmental Protection Agencies. Ü.V. was supported by the European Community's 6th Framework Marie Curie Intra-European Fellowships programme.

References

Andolfatto P (2005) Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*, **437**, 1149–1152.

Aspi J, Roininen E, Ruokonen M, Kojola I, Vila C (2006) Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population. *Molecular Ecology*, **15**, 1561–1576.

Barbara T, Palma-Silva C, Paggi GM *et al*. (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Molecular Ecology*, **16**, 3759–3767.

Bell GI, Jurka J (1997) The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *Journal of Molecular Evolution*, **44**, 414–421.

Brandström M, Ellegren H (2008) A genome-wide analysis of microsatellite polymorphism circumventing the ascertainment bias. *Genome Research*, **18**, 881–887.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**, 249–256.

Carroll SB (2005) Evolution at two levels: on genes and form. *Public Library of Science, Biology*, **3**, 1159–1166.

Cegelski CC, Waits LP, Anderson NJ *et al*. (2006) Genetic diversity and population structure of wolverine (*Gulo gulo*) populations at the southern edge of their current distribution in North America with implications for genetic viability. *Conservation Genetics*, **7**, 197–211.

Chamary JV, Hurst LD (2004) Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage. *Molecular Biology and Evolution*, **21**, 1014–1023.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, **15**, 1496–1502.

Ellegren H (1991) DNA typing of museum birds. *Nature*, **354**, 113–113.

Ellegren H (1992) Polymerase-chain-reaction (PCR) analysis of microsatellites — a new approach to studies of genetic relationships in birds. *Auk*, **109**, 886–895.

Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, **24**, 400–402.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite evolution — directionality or bias. *Nature Genetics*, **11**, 360–362.

Ellegren H, Moore S, Robinson N *et al*. (1997) Microsatellite evolution — a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution*, **14**, 854–860.

Ellegren H, Smith NGC, Webster MT (2003) Mutation rate variation in the mammalian genome. *Current Opinion in Genetics and Development*, **13**, 562–568.

Flagstad O, Walker CW, Vila C *et al*. (2003) Two centuries of the Scandinavian wolf population: patterns of genetic variability and migration during an era of dramatic decline. *Molecular Ecology*, **12**, 869–880.

Flagstad O, Hedmark E, Landa A *et al*. (2004) Colonization history and noninvasive monitoring of a reestablished wolverine population. *Conservation Biology*, **18**, 676–688.

Frankel OH, Soulé ME (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.

Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**, 2467–2474.

Hellborg L, Walker CW, Rueness EK *et al*. (2002) Differentiation and levels of genetic variation in northern European lynx (Lynx lynx) populations revealed by microsatellites and mitochondrial DNA analysis. *Conservation Genetics*, **3**, 97–111.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, **8**, 3–17.

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics*, **22**, 253–259.

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics*, **13**, 74–78.

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences, USA*, **95**, 10774–10778.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455–1460.

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. *Molecular Biology and Evolution*, **21**, 991–1007.

Liberg O, Andrén H, Pedersen H-C, Sand H, Sejberg D, Wabakken P, Åkesson M, Bensen S (2005) Severe inbreeding depression in a wild wolf (*Canis lupus*) population. *Biology Letters*, **1**, 17–20.

Linnell JV, Salvatori V, Boitani L (2007) *Guidelines for Population Level Management Plans for Large Carnivores*. Prepared by Large Carnivore Initiative for Europe, IUCN/SSC Working Group. Istituto di Ecologia Applicata, Rome, Italy.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R (2004) Population genetic analysis of ascertained SNP data. *Human Genomics*, **1**, 218–224.

Pardi F, Sibly RM, Wilkinson MJ, Whittaker JC (2005) On the structural differences between markers and genomic AC microsatellites. *Journal of Molecular Evolution*, **60**, 688–693.

Park SDE (2001) *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*, University of Dublin, http://acer.gen.tcd.ie/~sdepark/ms-toolkit/index.php.

Payseur BA, Cutter AD (2006) Integrating patterns of polymorphism at SNPs and STRs. *Trends in Genetics*, **22**, 424–429.

Petes TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics*, **146**, 491–498.

Pheasant M, Mattick JS (2007) Raising the estimate of functional human sequences. *Genome Research*, **17**, 1245–1253.

Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847–859.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Reed DH, Lowe EH, Briscoe DA, Frankham R (2003) Fitness and adaptation in a novel environment: effect of inbreeding, prior environment, and lineage. *Evolution*, **57**, 1822–1828.

Rolfsmeier ML, Lahue RS (2000) Stabilizing effects of interruptions on trinucleotide repeat expansions in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **20**, 173–180.

Rozas J, Sanchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity*, **98**, 692–704.

Santibanez-Koref MF, Gangeswaran R, Hancock JM (2001) A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Molecular Biology and Evolution*, **18**, 2119–2123.

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, **20**, 211–215.

Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature*, **354**, 63–65.

Schug MD, Mackay TFC, Aquadro CF (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genetics*, **15**, 99–102.

Schug MD, Hutter CM, Noor MAF, Aquadro CF (1998) Mutation and evolution of microsatellites in *Drosophila melanogaster*. *Genetica*, **103**, 359–367.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–629.

Sibly RM, Meade A, Boxall N *et al*. (2003) The structure of interrupted human AC microsatellites. *Molecular Biology and Evolution*, **20**, 453–459.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, **76**, 449–462.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

The ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.

Vilà 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.

Walker CW, Vila C, Landa A, Linden M, Ellegren H (2001) Genetic variation and population structure in Scandinavian wolverine (*Gulo gulo*) populations. *Molecular Ecology*, **10**, 53–63.

Waterson GA (1975) Number of segregating sites in genetic models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Weber JL (1990) Informativeness of human (dC-dA)n(dG-dT)n polymorphisms. *Genomics*, **7**, 524–530.

Wiehe T (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theoretical Population Biology*, **53**, 272–283.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.

Ülo Väli is a Marie Curie postdoc interested in conservation genetics of birds and mammals. Research in the Ellegren laboratory integrates genomics and evolutionary biology.

Supplementary material

The following supplementary material is available for this article:

Table S1 PCR details for intron markers. Intron lengths are given separately for wolves and coyotes[1], wolverines[2] and lynx[3]

Table S2 Microsatellite markers used. N is the number of markers

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03876.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.

© 2008 The Authors
Journal compilation © 2008 Blackwell Publishing Ltd