

Genomics and conservation genetics

Michael H. Kohn¹, William J. Murphy², Elaine A. Ostrander³ and Robert K. Wayne⁴

¹ Department of Ecology & Evolutionary Biology, Rice University, MS 170, 6100 Main Street, Houston, TX 77005, USA

² Department of Veterinary Integrative Biosciences, Texas A&M University, MS 4458, College Station, TX 77843-4458, USA

³ National Human Genome Research Institute, National Institute of Health, Comparative Cancer Section, 50 South Drive, MSC 8000, Bethesda, MD 20892-8000, USA

⁴ Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095-1606, USA

In large part, the relevance of genetics to conservation rests on the premise that neutral marker variation in populations reflects levels of detrimental and adaptive genetic variation. Despite its prominence, this tenet has been difficult to evaluate, until now. As we discuss here, genome sequence information and new technological and bioinformatics platforms now enable comprehensive surveys of neutral variation and more direct inferences of detrimental and adaptive variation in species with sequenced genomes and in 'genome-enabled' endangered taxa. Moreover, conservation schemes could begin to consider specific pathological genetic variants. A new conservation genetic agenda would utilize data from enhanced surveys of genomic variation in endangered species to better manage functional genetic variation.

Introduction

Given the current and future capacity of genome centers, hundreds of species of animal, plant, fungi, bacteria and viruses will eventually have their genome sequenced ([1]; c.f. GOLDTM, the Genomes OnLine Database v 2.0 at <http://www.genomesonline.org/>). Additionally, thousands of previously unknown microbial species will emerge from so-called 'metagenomic' sequencing projects [1]. As of 2006, more than 50 plant and 215 animal genome projects are ongoing, covering at varying resolution >35 vascular plant species, 135 invertebrates including sponges, echinoderms, cnidarians and insects, and 75 vertebrate species. Close to 1000 red-listed mammal species alone are within the same order or group to that containing a genome-sequenced species. Most genera represented by genome-sequenced species also contain endangered or vulnerable species, subspecies, or populations (Table 1). Thus, depending on phylogenetic distance and the questions being addressed, the sequencing of one species, such as the dog *Canis familiaris* will result in numerous 'genome-enabled' taxa (e.g. Figure 1) [2,3].

In addition to demographic and environmental uncertainties, genetic uncertainties (see Glossary) such as founder effects, genetic drift and inbreeding, threaten endangered species [4–11]. Conservation genetics addresses these issues and numerous others with limited surveys of neutral marker variation (Box 1). The results of these studies have a well-established role in conservation and

have often aided the management of rare and endangered species. Consequently, given limited resources, there might be little incentive to move from conservation genetics to 'conservation genomics', particularly when traditional conservation genetic issues in vertebrate animals and vascular plants are considered (Box 1).

So which of the more traditional conservation genetics applications, if any, are in need of genomic tools and information? Clearly, many applications of molecular markers in conservation could be transformed by genomic approaches [12–15], and approaches such as metagenomics could give rise to entirely novel conservation research questions. However, one of the most immediate questions, in our view, is how genomic resources might be used to test the implicit assumption of conservation genetics, that the observed levels of neutral genetic variation can be used to predict the levels of detrimental variation accrued and adaptive variation lost by endangered species owing to population size decline [4,16]. Existing data suggest that the correlation of neutral with detrimental and adaptive variation is low, in large part, because this relationship is contingent upon genomic sampling and population-specific demographic history (Table 2).

This concern suggests three research areas where genomics could enhance traditional small-scale surveys of neutral marker variation. First, expanded neutral marker surveys enable the more detailed reconstruction of the demographic histories of species [17], thereby facilitating more precise predictions regarding the increase and loss, respectively, of detrimental and adaptive genetic variation (Table 2). Second, expanded DNA sequence polymorphism surveys combined with bioinformatical analyses (Box 2) [18] could be used to quantify levels of detrimental genetic variation that might diminish fitness and threaten population persistence and of adaptive variation, which enables a genetic response to environmental change [11]. Understanding the relationship of neutral variation with detrimental and adaptive variation is essential for the effective use of genetic data in conservation [4]. Finally, information about neutral and adaptive variation could be used to preserve population distinctiveness within a framework of genetic and ecological exchangeability (*sensu* Templeton) [19–21].

Genetic variation considered in conservation

Each of the three types of genetic variation in nature (neutral, detrimental and adaptive) has specific uses and conceptual implications for conservation.

Corresponding author: Kohn, M.H. (hmkohn@rice.edu).
Available online 5 September 2006.

Glossary

Adaptive genetic variation: genetic variation that affects fitness. A population might be fixed for an adaptive variant or polymorphic because the variant is newly arisen or is maintained by balancing selection.

Comparative genomics: evolutionary relationship between the genes and proteins of different species, and applications that draw upon these relationships to infer structure and function.

Demographic uncertainty: uncertainty of population persistence resulting from the effects of random events on the survival and reproduction of individuals, such as extremely skewed sex ratio or threshold effects (e.g. Allee effect).

Detrimental genetic variation: variation that has a negative effect on fitness, often brought into the population by mutation or gene flow and sometimes increased by drift; can be strongly detrimental (recessive) mutations that can cause inbreeding depression in the homozygous state, or weakly detrimental mutations whose aggregate effect reduces population fitness as genetic load.

Ecological exchangeability: factors that define the fundamental niche and the limits of spread of new genetic variants through genetic drift and natural selection. Exchangeability is rejected when there is evidence for population differentiation owing to genetic drift or natural selection.

Environmental uncertainty: unpredictable events (e.g. variation in food supply or exposure to parasites or competitors) that adversely affect the long-term survival of the population as a whole.

Functional variation: characterization of genes and proteins with respect to function. Examples of functional properties include the structure of the resulting protein or the levels of the expression of the gene. Functional properties can be predicted (e.g. by reconstructing the 3D structure of proteins based on the gene sequence, or by predicting gene expression based on regulatory sequence motifs on the promoter). Finally, other approaches merely infer function (e.g. based on the conservation of gene sequences between species, the type and location of a SNP, or the signature of natural selection on a gene sequence).

Gene annotation: database entries that describe what the gene or genomic region is, provide its boundaries and functional elements (e.g. transcription factor binding sites), background information about the expression of the gene and disease-causing mutations, known alleles, and cross-references to other databases.

Genetic exchangeability: factors that define the limits of spread of new genetic variants through gene flow. Exchangeability is rejected when there is evidence of restricted gene flow between populations.

Genetic uncertainty: stochastic factors (e.g. founder effects, inbreeding and genetic drift) that predominantly determine the amount and spatial distribution of neutral, detrimental and adaptive genetic variation in small populations.

Genome-enabled endangered taxa: many of the resources yielded by genome projects (e.g. genomic libraries, DNA and RNA arrays, bioinformatics tools and databases), have cross-species applicability, such that genome-sequenced species result in 'genome-enabled' taxa, including rare and endangered ones. Resource portability varies with phylogenetic distance and the questions being addressed. Genome-enabling is not only characterized by the availability of novel tools, resources and more data, but also by a transition from bioinformatics or statistical analyses of a few loci to powerful genome-wide inferences and the routine use of databases.

Neutral variation: variation that has a small selection coefficient relative to the population size, such that $|s| < 1/2N_e$, where s is either the selective disadvantage of a detrimental genetic variant or the selective advantage of an adaptive genetic variant and N_e is the effective population size.

Neutral variation

Neutral genetic variation in natural populations is predominantly governed by the interaction of mutation, genetic drift, recombination and migration [22]. Whereas genetic drift results in the stochastic loss of genetic variants, mutation, recombination and migration introduce new variants or generate novel combinations of existing ones. Important conservation genetic issues rely on information from small-scale neutral marker surveys, including inferences of recent and historical gene flow useful to evaluate distinctiveness and conservation status of populations (Box 1). Low levels of neutral variation typically raise concerns regarding inbreeding depression, genetic load and loss of adaptive potential (Table 2) [4,16].

However, although variation in a few neutral markers might conceptually predict levels of detrimental and adaptive variation, empirical studies suggest that neutral variation does not correlate closely with detrimental

and adaptive variation and underestimates population differentiation in quantitative traits [4,20,23]. Hence, limited neutral marker surveys might result in misconceptions about levels of detrimental and adaptive variation in endangered species.

Detrimental variation

Among the genetic variants that can adversely affect fitness are those that alter the protein sequence of genes (Box 2) [24,25]. The dynamics of such mutations varies: selection is most efficient in large populations and, consequently, even detrimental mutations of small effect are eliminated or kept at low frequencies such that the genetic load is low [26–29]. In small populations, weakly selected mutations, in the range of $-1/2N < s < 1/2N$ (where N is the genetically effective population size and s the selection coefficient) behave as although they are neutral and therefore can increase in frequency owing to drift. This increase in small populations lowers the mean fitness and, even in a constant environment, can lead to a 'mutation meltdown' [6,7].

Inbreeding depression caused by increased probability of homozygosity for detrimental alleles is a more immediate concern [4,30]. Under some demographic scenarios, strongly detrimental mutations are likely to be purged during population-size declines (Table 2) [4,16]. For example, populations that have long persisted at a small size might be less prone to inbreeding depression than are those that have declined over short time periods.

To reduce genetic load and to minimize the probability of inbreeding depression, populations are managed such that levels of detrimental variation should be low [31]. In captive populations, breeding strategies can successfully minimize the loss of genetic variation while increasing the chance of purging disease alleles, as was done for hereditary blindness in wolves *Canis lupus* [32]. In the wild, strategies to increase breeding population size and movement between populations can lower the genetic load and risk of inbreeding depression. For example, the translocation of Texas cougars *Felis concolor stanleyana* to the endangered Florida cougar *F. c. coryi* population reduced the frequency of detrimental phenotypic effects in the Florida population.

Adaptive variation

Variation in genes enables populations to respond to environmental challenges, such as climate change or exposure to a new pathogen [8,11]. For several reasons, population size reductions are predicted to result in the concomitant reduction of adaptive genetic variation (Table 2). First, less standing genetic variation is sustained in small populations [4]. Second, fewer new mutations per unit time appear in small populations and the opportunity for recombination is reduced. Even in large populations, beneficial new mutations rarely occur [22]. Third, for the same reasons that selection fails to remove weakly detrimental mutations, the probability that an adaptive mutation of small effect can be maintained by selection is reduced [4]. Consequently, the potential of endangered species to adapt by means of weakly adaptive alleles is limited.

Maintaining adaptive genetic variation is an important conservation strategy [11,19,20], which requires

Table 1. Genome-sequenced mammals and examples of genome-enabled endangered taxa

Order/Group	Genome-sequenced species ^a				IUCN listed species ^d	Genome-enabled IUCN listed taxa ^e
	High coverage	Draft assembly ^b	Low coverage (~2X)	Other ^c		
Carnivora	Dog <i>Canis familiaris</i> **	Cat <i>Felis catus</i> *		Wolf <i>C. lupus</i> , coyote <i>C. latrans</i>	79	Ethiopian wolf <i>C. simensis</i> , Mexican wolf <i>C. l. baileyi</i> , black-footed cat <i>F. nigripes</i> , Cape pangolin <i>M. temminckii</i>
Pholidota			Chinese pangolin <i>Manis pentadactyla</i> **		0	
Perissodactyla			Horse <i>Equus caballus</i> *		12	Przewalski's horse <i>E. c. przewalski</i> , Grevy's zebra <i>E. grevyi</i> , African ass <i>E. africanus</i>
Cetartiodactyla		Cow <i>B. taurus</i> ***	Llama <i>Lama</i> sp.*, bottle-nosed dolphin <i>Tursiops truncatus</i> *	Zebu <i>Bos indicus</i> , Sheep <i>Ovis aries</i> , pig <i>Sus scrofa</i> , false killer whale <i>Pseudorca crassidens</i>	88	Banteng <i>B. javanicus</i> , guanaco <i>L. guanicoe</i> , argali <i>O. ammon</i> , Indian bottle-nosed dolphin
Chiroptera			Little brown bat <i>Myotis lucifugus</i> *, microbat <i>Microchiroptera</i> spp.*, megabat <i>Cynopterus</i> sp.**		248	Gray bat <i>M. grisescens</i> , fruit bat <i>Megachiroptera</i> spp.
Eulipotyphla			European common shrew <i>Sorex araneus</i> ***, European hedgehog <i>Erinaceus europaeus</i> **, mole <i>Cryptomys</i> sp.*	Four-toed hedgehog <i>Atelerix albiventris</i>	6	St. Lawrence Island shrew <i>S. jacksoni</i> , mole rat <i>Heterocephalus glaber</i>
Rodentia	Mouse <i>Mus musculus</i> ***, rat <i>Rattus norvegicus</i> ***	Guinea pig <i>Cavia porcellus</i> *	Ground squirrel <i>Spermophilus tridecemlineatus</i> **, kangaroo rat <i>Dipodomys</i> sp.*		315	Servant mouse <i>M. famulus</i> , summit rat <i>R. baluensis</i> , giant kangaroo rat <i>D. ingens</i>
Lagomorpha		Rabbit <i>Oryctolagus cuniculus</i> *	Pika <i>Ochotona</i> sp.*		17	Helan Shan pika <i>O. argentata</i>
Primates	Human <i>Homo sapiens</i> ***	Chimpanzee <i>Pan troglodytes</i> **, orangutan <i>Pongo pygmaeus</i> **, gibbon <i>Nomascus leucogenys</i> *, macaque <i>Macaca mulatta</i> **, marmoset <i>Callithrix jacchus</i> **	Mouse lemur <i>Microcebus murinus</i> *, tarsier <i>Tarsius</i> sp.*, bushbaby <i>Otolemur garnetti</i> **	Ring-tailed lemur <i>Lemur catta</i> , baboon <i>Papio anubis</i> and <i>P. hamadryas</i>	114	Bonobo <i>P. paniscus</i> , Mentawai macaque <i>M. pagensis</i> , buffy-headed marmoset <i>C. flaviceps</i> , golden-brown mouse lemur <i>M. ravelobensis</i>
Dermoptera			Flying lemur <i>Dermoptera</i> sp.*		1	Philippine flying lemur <i>Cynocephalus volans</i>
Scandentia		Tree shrew <i>Tupaia belangeri</i> *			6	Bornean tree shrew <i>Tupaia longipes</i>
Xenarthra		Nine-banded armadillo <i>Dasypus novemcinctus</i> **	Hoffmann's two-toed sloth <i>Choloepus hoffmanni</i> *		6	Hairy long-nosed armadillo <i>D. pilosus</i>
Proboscidea		African savanna elephant <i>Loxodonta africana</i> *		Asian elephant <i>Elephas maximus</i>	2	
Hyracoidea			Rock hyrax <i>Procavia capensis</i> *		0	Hoggar hyrax <i>Heterohyrax antineae</i>
Macroscelidea			Elephant shrew <i>Elephantulus</i> sp.*		2	
Afrosoricida			Lesser hedgehog tenrec <i>Echinops telfairi</i> ***	Tailless tenrec <i>Tenrec ecaudatus</i>	16	
Marsupials		Opossum <i>Monodelphis domestica</i> **	Tammar wallaby <i>Macropus eugenii</i> **		55	Pygmy short-tailed opossum <i>M. kunsii</i> , Barrow Island euro <i>M. robustus isabellinus</i>
Monotremes		Platypus <i>Ornithorhynchus anatinus</i> **			1	

^aSequencing goal (from high to low, c.f. <http://www.genome.gov/10002154> and links to genome centers and genome projects); progress as of November 2005: *** completed,

** in process, or * in pipeline (c.f. <http://www.genome.gov/10002154>).

^bWhen draft assembly is in the pipeline, low coverage sequencing is usually in progress or complete.

^cUnspecified sequencing efforts are likely to include, for example, cDNA sequencing (e.g. zebu) and large-scale SNP discovery (e.g. wolf and coyote), c.f. GOLDTM, the Genomes OnLine Database v 2.0 at <http://www.genomesonline.org/> [1].

^dSpecies listed in IUCN red list categories extinct–vulnerable (<http://www.redlist.org/>).

^eExamples of species, subspecies, or subpopulation belonging to the same genus as genome-sequenced species that are listed in all IUCN red list categories (excluding least concern).

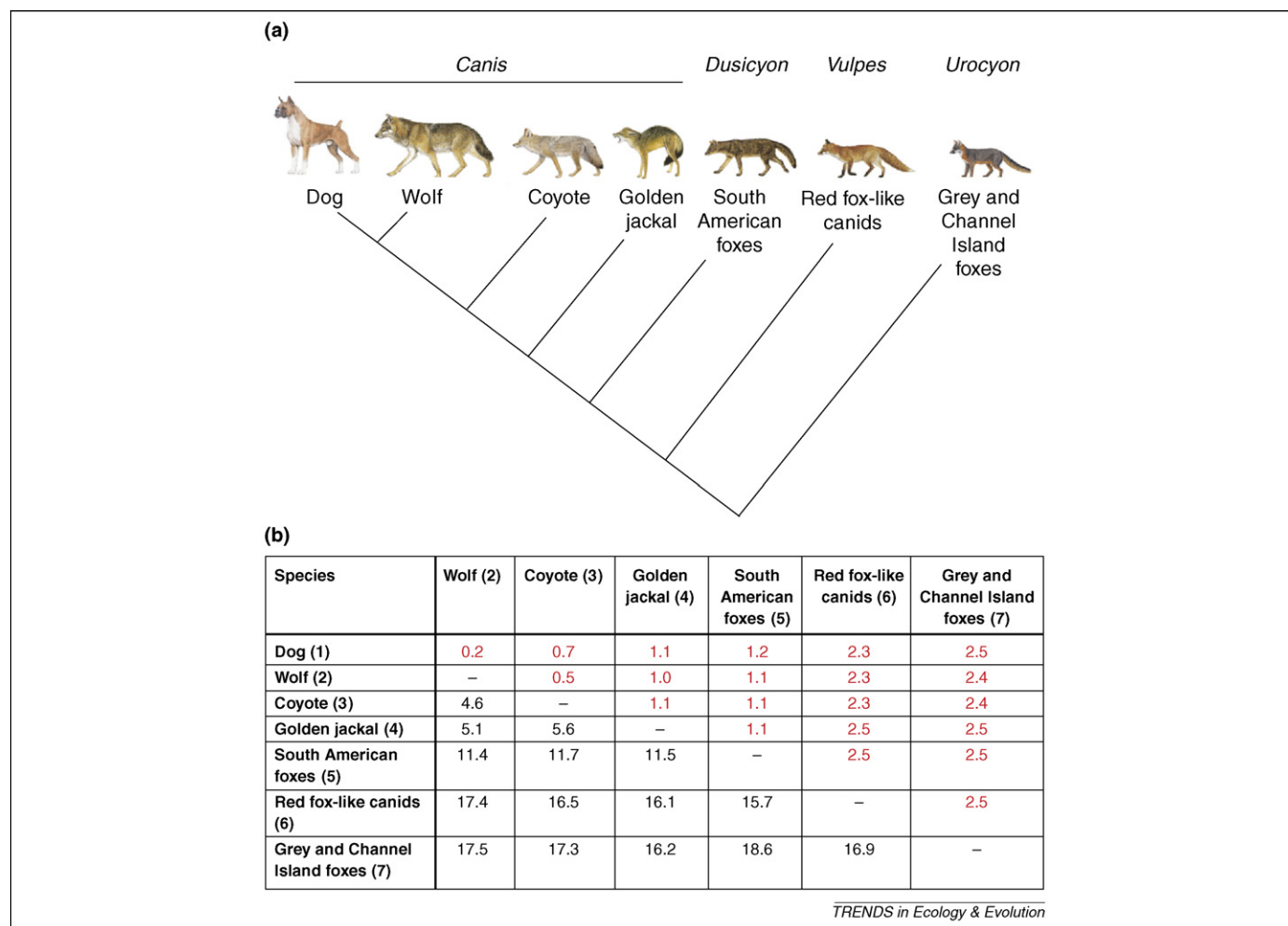


Figure 1. Taxonomic relationships (a) and percentage genetic divergence (b) between canids potentially genome enabled by the 1.5× and 7.5× coverage genome sequences of poodle and boxer, respectively. Based on intron (in red) [2] and mitochondrial DNA (in black) sequences (c.f. [80]). Genomic platforms, such as microarrays, have proven to be portable between human and chimpanzee (nuclear sequence divergence ~1%) and human and macaque (nuclear sequence divergence ~3%).

identification of populations that show local genetic adaptations, and other evidence for adaptive divergence, such as in life-history traits and ecology [33]. Thus, subsets of populations that maximally represent adaptive variation in an endangered species could hypothetically be targeted for conservation [19,33].

Functional genomic measures of endangerment

High-throughput genotyping technologies and easier access to markers enable broader-scale surveys of genomic variation in endangered species and detection of variation at previously unstudied levels, such as gene expression. Here, we define research agendas that enable tests of the

Table 2. Predicting detrimental and adaptive genetic variation based on observed neutral marker variation^a

Demographic scenario		Observed neutral variation	Predicted detrimental variation	Predicted adaptive variation
Equilibrium	Large population	High	High	High
	Small population	Low	Low	Low-medium (retention more than neutrality)
Bottleneck	Shortly after	Low (loss of alleles more than heterozygosity)	Low (loss of lethals more than detrimental)	Low-medium (retention more than neutrality, dependent on mutation rates)
	Some time after	Low-high (depending upon mutation rates)	High (assuming high mutation rates)	Low-medium (retention more than neutrality)
Metapopulation	Mixture of separately bottlenecked populations	High (owing to fixation of different variants in each population)	Low (owing to purging of different variants in each population)	Low-medium
	Extinction recolonization dynamics and low gene flow	Low (owing to low effective population size)	Low (owing to purging)	Low-medium (depending on selection model)
	No extinction and substantial gene flow	High (as if one large population)	High (as if one large population)	High (as if one large population)

^aAdapted with permission from [4].

Box 1. Neutral marker applications in conservation

The use of limited neutral marker surveys to measure genetic variation is routinely done for rare or endangered species [15,35,66]. Molecular markers that are assumed to be selectively neutral include mini- and microsatellites, mitochondrial DNA, chloroplast DNA, restriction length polymorphisms, AFLPs and SNPs [15,35,66]. Adherence to neutrality and the merits of marker uses have been discussed elsewhere [15,35,66,67].

Despite its long history, the debate surrounding the relevance of neutral genetic variation to the fitness of individuals and endangered populations remains controversial. In part, the lack of conclusive evidence for a strong relationship between heterozygosity, inbreeding and fitness could be attributed to the fact that only a small fraction of the genome is sampled and that neutral marker variation is affected primarily by drift and mutation whereas functional genes are additionally influenced by selection (e.g. [4,68]). Moreover, only recently are conservation genetic studies paying closer attention to reconstructing demographic history, population size estimation and the detection of past founder events and bottlenecks. Parentage and kinship analyses are often set within the context of conservation and can involve direct and indirect estimates of migration rates (e.g. [69]).

With respect to the definition of populations and species that deserve particular attention for conservation, studies of population structure and intraspecific phylogeny are abundant in the literature, and their interpretations have led to discussion regarding the units for conservation and management [19]. Often, the identity of stocks and the geographical origin of individuals are of practical relevance (e.g. in wildlife forensics) [14] as is the use of non-invasively collected genetic data for independent assessments of population size and social structure [70]. At higher taxonomic levels, phylogenetic approaches have been used to clarify species status or to quantify biological diversity [19]. Other issues related to species status concern the detection of invasive species, or the detection of hybridization between endangered and non-endangered taxa [4].

predicted relationship between neutral and functional genetic variation by measuring the previously unknown quantities of detrimental genetic variation, strongly pathological variants and adaptive distinctiveness.

Expanding neutral marker surveys

Large-scale surveys can now be designed that use an order of magnitude more markers than previous studies (usually <15). Markers can be identified with ever-greater ease for the species whose genome sequence have been determined and for genome-enabled species. Databases provide tools that help establish gene orthology, gene function and genetic map locations of other markers and genes. New analytical approaches facilitate the study of single nucleotide polymorphisms (SNPs, Box 2) in populations [34,35] and provide an alternative to maternally inherited mitochondrial markers and microsatellite loci whose analysis is fraught with difficulties. An increase in the number of SNPs could compensate for their lower levels of polymorphism compared with microsatellites [15], as shown by a study of an inbred wolf population in Scandinavia [36].

Three specific pressing conservation issues could be resolved by expanding neutral marker surveys. First, more extensive measurements of variation are required to assess the value of neutral marker surveys to predict levels of detrimental and adaptive variation. Second, reconstruction of the demographics of endangered species is necessary to distinguish long-term evolutionary effects from recent human-induced changes. For example, in the orangutan *Pongo pygmaeus*, genetic analysis of populations on

Box 2. Predicting detrimental SNPs

SNPs occur throughout animal and plant genomes [67]. Only ~1–3% of a typical genome is protein encoding and, within these regions, non-synonymous SNPs can be considered candidates for functional change [24,48,51,52]. The phenotypic effect of any particular SNP is rarely known and often can only be inferred based on the evolutionary dynamics of the variant or on its effect on protein function [22,47,48]. The nonsynonymous:synonymous SNPs ratio can be taken as a measure of the strength of purifying selection on a gene or, when summed over many genes, the entire genome [22,47]. More refined measures rank SNP effects from benign to radical and consider the effect of SNPs on protein structure, function, cellular localization and differences in the biochemical and structural properties of amino acids [22,24,25,42,48,52,71]. The human SNPEffect database (<http://snpeffect.vib.be/>) predicts the consequences of SNPs on proteins [24]. Some genes, once mutated, appear more prone to causing disease than do others (e.g. those genes detailed in the Disease Gene Conserved Sequence Tag database DG-CST) [52].

Even synonymous SNPs in protein-encoding genes could have functional implications. Although multiple codons can encode the same amino acid, some occur more commonly in the genome than is predicted by random (i.e. codon usage bias) [72]. Therefore, a SNP that causes a change from a more common or preferred codon to a rare or unpreferred codon can affect the efficiency of protein synthesis and expression.

Most SNPs occur in the non-encoding portion of the genome, but can nevertheless be evaluated with regard to function [73]. A substantial fraction of the non-encoding genome is conserved between species, suggesting that purifying selection acts on a large portion of the genome. Thus, SNPs can be evaluated based on their location in conserved versus non-conserved non-encoding regions. Moreover, the regulatory regions of genes (e.g. promoters and enhancers) have been annotated using comparative and predictive algorithms (e.g. [73]), thereby enabling the assessment of non-encoding regulatory SNPs. For instance, SNPs that occur in the transcription factor binding sites of a promoter are more likely to affect function (i.e. gene regulation) than are SNPs that occur outside the regulatory region of a gene.

Borneo showed demographic effects owing to recent rather than long-historic habitat loss and suggested that the population was at risk of inbreeding depression and increased genetic load [37]. By contrast, the recent demographic crash of the Australian northern hairy-nosed wombat *Lasiorhinus krefftii* has not caused a genetic decline as profound as that due to long-term trends [38]. Clearly, these and other high-profile cases indicate that we should perhaps not consider endangered species as all being equally threatened by genetic factors but, instead, that we could prioritize them based on their particular demographic histories and predicted genetic threats to their long-term persistence (Table 2). Finally, the detection of adaptation at 'outlier loci' during genome scans is accomplished best when baseline levels of neutral genetic variation across the genome have been determined [39–42].

Detecting weakly detrimental genetic variation

In small and declining populations, weakly detrimental genetic variants could rise to high frequency owing to drift [4,16,29]. Although of little consequence individually, it is the aggregate effect of weakly detrimental mutations that could be harmful to endangered species. Relaxation of purifying selection owing to population size reduction can be detected at the DNA sequence level [22]. Unpreferred synonymous mutations or nonsynonymous

mutations (Box 2) affect protein function and usually occur at low frequency in populations, suggesting that they are weakly detrimental [43,44].

Contrasts between large and small populations can reveal the altered dynamics of detrimental genetic variants. For example, *Drosophila madeirensis*, which is endemic to Madeira, and *D. guanche*, an insular Canary Island species, display higher rates of substitutions to unpreferred synonymous codons than expected based on comparisons with abundant mainland species [43]. Similar effects of small population size are found in Hawaiian drosophilids and in the *D. simulans* species complex [28,29]. A study comparing island species of dabbling ducks and doves with their related mainland counterparts failed to find higher rates of substitutions to unpreferred codons, but found an elevated rate of nonsynonymous compared with synonymous substitutions [44]. Similar to genome-wide studies that have contrasted species with large and small population sizes [45], conservation genetic surveys covering greater portions of the genome could provide powerful assays of the reduced efficiency of selection in endangered species compared with their non-endangered counterparts. Such information should be used to identify populations with high levels of detrimental variation and support specific actions, such as enhanced gene flow [46], that might increase the efficacy of selection.

The analysis of polymorphism with bioinformatics tools enables predictions of the deleterious effects of synonymous and nonsynonymous SNPs on proteins (Box 2). For example, an analysis of >31 600 human SNPs showed that only ~290 (>0.01%) change the subcellular localization of proteins [24]. These SNPs are rarely found in human populations, indicating that they are detrimental. By contrast, changes in protein stability are more frequent (>50% of all human SNPs), indicating that they are comparatively benign [24]. Consequently, one approach to link functional genomic changes with fitness or health is to develop a panel of several hundred to several thousand SNPs in exons and to contrast SNP frequencies in regions where they are predicted to be detrimental with regions where they are predicted to be benign. Populations with relatively elevated SNP levels at sites for which bioinformatics predicted detrimental consequences would be considered genetically most at risk.

A comparative genomics perspective can enhance the value of SNP surveys for conservation. Because deleterious polymorphism can persist at low frequency in populations but divergence between species is neutral or adaptive, a contrast of levels of polymorphism with divergence enables detrimental effects to be inferred [22,47]. For example, whereas ~35% of 551 disease-causing SNPs in humans map to sites that alter intrinsic structural features of proteins, only 9% of 225 differences between humans and chimps do, suggesting that such mutations are detrimental [48]. The historical effective population size of humans is surprisingly low (~15 000), and the observation of a substantial mutational load inferred from the abundance of weakly detrimental SNPs has led to speculations regarding public health consequences and senescence [26,48] (Box 2). Similar large-scale comparative contrasts of endangered with non-endangered species could reveal

the reduced efficiency of selection and its relationship to population demography.

Detecting strongly detrimental or pathological mutations

Knowledge about specific mutations that cause disease could inform the genetic management of small wild or captive populations [4,5,31,32]. For genome-enabled relatives of the soon-to-be-sequenced domestic cat *Felis catus*, such as the tiger *Panthera tigris* and cheetah *Acinonyx jubatus*, the comparative mapping of strabism and amyloidosis should already be feasible. Similarly, goiter is a genetic disease that affects the health of cattle *Bos taurus* and of several potentially genome-enabled species, including sheep *Ovis aries*, goats *Capra aegagrus*, buffalo *Bubalus arnee* and bongo antelope *Tragaelaphus eurycerus*. Lists of genetic diseases to be surveyed can be compiled from online databases, such as the Online Database for the Inheritance in Man (<http://www.ncbi.nlm.nih.gov/entrez/>) and the Online Database for Inheritance in Animals (<http://www.angis.org.au/omia/>) [49], a growing database currently covering 135 domesticated and wildlife species and >15 000 publications describing >2500 inherited disorders [49]. For dogs, 500 inherited disorders are on record in the 'Listing of Inherited Disorders in Animals' database (<http://www.vetsci.usyd.edu.au/lida/>) [50]. Conceivably, several disorders might affect the dog as well as wild canids. However, in general, these databases are in the early stages of development, and the work involved to map diseases is cumbersome, such that for most genome-enabled taxa, only a few specific genetic disorders can now be screened for.

More encompassing approaches to screen for potentially harmful variants in endangered species could rely on bioinformatic analyses of SNP data, because genes with predicted detrimental effects once mutated (i.e. morbidity genes) appear to share properties that can be detected by using bioinformatic tools [51]. For example, bioinformatic interrogation of >1000 human genes that are known for their detrimental effect on viability, fertility and longevity revealed shared characteristics, including a slow evolutionary rate, a wide phylogenetic distribution and large introns with precise splicing signals. Potentially, ~30% of all human genes could be morbidity genes, and >80% of these await confirmation based on clinical phenotypes [51]. Hence, SNP surveys seeking to infer the accumulation of strongly detrimental variants could target known or predicted morbidity genes to monitor for potentially harmful segregating variants in captive or managed populations of endangered species.

Bioinformatic tools facilitate the design of marker surveys aimed at the detection of strongly detrimental variants. For example, regions within genes that are known to contribute to disease once mutated are compiled in the Disease Gene Conserved Sequence Tags database (DG-CST, <http://dgcst.ceinge.unina.it/>) [52] (Box 3). Evolutionary conservation suggests functional importance and, thus, nucleotide changes in conserved regions of a disease-related gene are likely to have detrimental effects [52]. Consequently, a time- and money-saving strategy would be to restrict searches for detrimental variants to regions of

Box 3. 'Population genomics' in conservation

Three related approaches are useful for identifying outlier loci that, owing to local selection, deviate from the genome-wide levels of genetic variation and differentiation [39–42]. The main goal of these population genomic approaches is to narrow the search for genes under selection to smaller genomic regions or smaller sets of genes that can be tested for the signatures of selection. An increasing number of human genetic studies [17,55], studies of model organisms [56,57] and studies of non-model organisms use these approaches [58–62]; however, these 'natural selection mapping' approaches have yet to be applied to endangered species.

Candidate gene approaches

Genes are chosen for study as probable targets of selection *a priori*, based on function, previous observations of unusual genetic variation, or their map location within regions that contain quantitative trait loci (QTL). A classic example is the major histocompatibility complex (MHC) [4,68]. Based on genome annotation data, it is now possible to identify the genes that might be involved in a particular physiological process, such as immune responses, and to target a subset for analysis. Similarly, other groups of genes annotated or known to function within an interesting context, such as drought tolerance in plants, could be chosen for analysis [61].

Selective sweep approaches

Selection for phenotypic traits might leave a genetic signature in the genome that enables their identification [39–42]. The power to detect a selective sweep is determined by the strength of selection, population size and recombination rates [74]. For example, genomic regions of *Drosophila melanogaster* have become the target of positive selection during the past few thousand years as a probable consequence of the colonization of new habitats outside of its original range in Africa [75,76]. However, selectively neutral demographic models, such as bottlenecks, that explain the data equally well must also be considered [77].

Linkage disequilibrium mapping approaches

LD mapping is the means of identifying genes underlying continuously varying traits, or QTLs, and involves finding markers that segregate with specific traits in pedigrees [42]. However, the traits need to be identified in advance and LD mapping requires cross-breeding designs, which limits its applicability to endangered species. LD mapping is conceptually related to the selective sweep and utilizes the principle that genome regions under selection will have marker loci that are inherited in a non-random fashion, that is, they are in LD [74,78,79]. LD mapping has been used to map the gene causing rats *Rattus norvegicus* to be resistant to the rodenticide warfarin [56]. Furthermore, a single or several diagnostic loci might mark the haplotype (e.g. TagSNPs), thereby reducing the genotyping efforts [61]. Finally, a poorly explored issue in conservation concerns the level of LD in endangered populations and how it diminishes their ability to purge deleterious alleles and respond to selection (e.g. [81]).

genes for which bioinformatic analyses with the computer programs SNPeffect [24] or DG-CST [52] (Box 2), for example, predict that SNPs or other types of mutations, such as insertions or deletions, are detrimental. Knowledge of variants predicted to be harmful could be used to devise breeding strategies that promote their purging from the population or, by preventing inbreeding, minimize their probability of homozygosity [31,32].

Detecting adaptive genetic variation

Methods to detect adaptive genetic variation utilize sequence polymorphism within and divergence between species, or both in combination [22]. Lists of genes that show evidence for adaptive evolution inferred from

divergence data alone (the rate of nonsynonymous divided by the rate of synonymous divergence) have been compiled in genomic databases (e.g. <http://www.sbc.su.se/~liberles/TAED2002/> [53]). Such data show that adaptive evolution is a widespread property of species and suggest that life-history or ecological characteristics that promote the continued process of adaptation need to be identified and preserved [11,19]. For instance, the evolutionary rate of the *SEMG2* gene in higher primates scales positively with female promiscuity [54]. Therefore, retaining natural social systems in captive and wild primates might be essential for preserving the process that drives adaptive evolution.

A group of conceptually related 'population genomic' approaches, including candidate gene, selective sweep and linkage disequilibrium (LD) mapping approaches (Box 3), show great potential for detecting adaptive differences between populations [39–42]. All three approaches detect a signature of natural selection on genetic variation relative to a null distribution predicted from theory or observed in the data. Geographical mapping of such adaptive genetic variants identified with these approaches can be used to assess adaptive distinctiveness of endangered populations within a framework of genetic and ecological exchangeability [19,21,33]. Adaptive genetic variation could prove essential for species persistence.

Currently, these tools are most readily implemented in genome-sequenced model organisms or genome-enabled species [55–57]. However, even for non-model species, population genomics surveys are conceivable. For example, to assess molecular genetic changes in the common frog *Rana temporaria* along an altitudinal gradient, 392 amplified length polymorphisms (AFLPs) were compared among populations from the Alps [58]. Eight candidate loci potentially involved in adaptation to altitude emerged from the survey. Similarly, a genomic scan implicated the albumin locus in local adaptive differences between high- and low-altitude deer mice *Peromyscus maniculatus* [59]. In such genome-wide surveys, commonly between 1–5% of loci emerge as candidates for local adaptation [58–62] which can provide sufficient power to identify adaptively divergent populations necessary for preservation of adaptive diversity and to evaluate the value of neutral variation to predict adaptive variation.

Detecting other sources of functional variation

The timing and localization of gene expression, which can be quantified by RNA microarrays, provides a new measure of genetic variation in natural populations that was previously hidden from view [42,63]. Gene regulation can differ in a systematic fashion across environments or along natural clines. For example, expression differences among killfish *Fundulus heteroclitus* within populations were found for ~18% of genes assayed, but expression differences among populations exceeded this number [64]. Moreover, expression of metabolic genes along a steep thermal gradient showed that, although much of the variation in gene expression was governed by drift, variation in expression for 22% of the genes followed the thermal cline in a fashion consistent with selection [64]. Conventional surveys of marker variation probably provide an incomplete

picture of this important source of genetic and quantitative variation [41,65]. Presumably, gene expression differences can be neutral, detrimental or adaptive, but if adaptive, they might identify ecologically unique populations.

A new agenda for conservation genetics

An immediate research program aimed at uncovering the empirical relationships between neutral, detrimental and adaptive genetic variation is needed. This should consist of studies that combine detailed demographic analyses based on neutral marker data with the application of approaches that infer the levels of detrimental and adaptive variation. Both endangered and model species should be included so that a wide variety of demographic and natural evolutionary experiments can be conducted. This would include the study of strongly detrimental variants, which is necessary to gauge their frequency in captive and wild genetically managed populations. Clearly, this effort represents a long-term research task, but it could lead to the expansion of databases of known genes (either predicted or observed) that could be monitored. Finally, efforts to identify genes that are subject to selection in natural populations should be initiated. This information should be used to develop rational schemes to preserve adaptive diversity within species.

The promise of conservation genomics is that developing databases and new molecular and analytical tools can be used to better infer the genetic health and historical and adaptive uniqueness of endangered populations. Paramount in this effort is the quantification of deleterious and adaptive genetic variation. Although considerable work is required to confirm experimentally the fitness consequences, either adaptive or detrimental, of genetic variants, bioinformatics and population genomic approaches can help circumvent this limitation. Furthermore, a comparative approach utilizing contrasts in variation between endangered and closely related non-endangered taxa could highlight changes in levels of detrimental or adaptive variation. Genomic resources and analytical approaches have so far been largely applied to model organisms but can now be applied to genome-enabled endangered taxa and, to a limited extent, exotic species. The toolkit of genomic science and a shift towards the study of functional genetic variation will improve our understanding of the pattern and process affecting genetic variation in rare and endangered species. However, new debates should be initiated concerning how this information can best be used to preserve and manage threatened populations and whether the expense in time and effort justifies the potential returns.

Acknowledgements

We thank Phil Hedrick, Jay Storz and two additional reviewers for providing insightful comments; Phil Morin for his reading of an earlier version of the article, and John Pollinger for computing genetic distances of the canids.

References

- Liolios, K. *et al.* (2006) The genomes on line database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res.* 34, D332–D334
- Lindblad-Toh, K. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819
- O'Brien, S.J. *et al.* (1997) Comparative genomics: lessons from cats. *Trends Genet.* 13, 393–399
- Hedrick, P.W. (2001) Conservation genetics: where are we now? *Trends Ecol. Evol.* 16, 629–636
- Zhang, Y.P. *et al.* (2002) Genetic diversity and conservation of endangered animal species. *Pure Appl. Chem.* 74, 575–584
- Gaggiotti, O.E. (2003) Genetic threats to population persistence. *Ann. Zool. Fenn.* 40, 155–168
- Spielman, D. *et al.* (2004) Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15261–15264
- Frankham, R. (2003) Genetics and conservation biology. *C. R. Biol.* 326, S22–S29
- DeSalle, R. (2005) Conservation genetics – genetics at the brink of extinction. *Heredity* 94, 386–387
- Amos, W. and Balmford, A. (2001) When does conservation genetics matter? *Heredity* 87, 257–265
- Lande, R. and Shannon, S. (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50, 434–437
- DeSalle, R. and Amato, G. (2004) The expansion of conservation genetics. *Nat. Rev. Genet.* 5, 702–712
- Hedrick, P.W. (2004) Recent developments in conservation genetics. *For. Ecol. Manage.* 197, 3–19
- Ryder, O.A. (2005) Conservation genomics: applying whole genome studies to species conservation efforts. *Cytogenet. Genome Res.* 108, 6–15
- Wayne, R.K. and Morin, P.A. (2004) Conservation genetics in the new molecular era. *Front. Ecol. Env.* 2, 89–97
- Kirkpatrick, M. and Jarne, P. (2000) The effects of a bottleneck on inbreeding depression and the genetic load. *Am. Nat.* 155, 154–167
- Akey, J.M. *et al.* (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2, 1591–1599
- Sunyaev, S. *et al.* (2001) Integration of genome data and protein structures: prediction of protein folds, protein interactions and 'molecular phenotypes' of single nucleotide polymorphisms. *Curr. Opin. Struct. Biol.* 11, 125–130
- Crandall, K.A. *et al.* (2000) Considering evolutionary processes in conservation biology. *Trends Ecol. Evol.* 15, 290–295
- Reed, D.H. and Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conserv. Biol.* 17, 230–237
- Moritz, C. (2002) Strategies to protect biological diversity and the processes that sustain it. *Syst. Biol.* 51, 238–254
- Fay, J.C. and Wu, C.I. (2003) Sequence divergence, functional constraint and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* 4, 213–235
- Merila, J. and Crnokrak, P. (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14, 892–903
- Reumers, J. *et al.* (2005) SNPeff: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. *Nucleic Acids Res.* 33, D527–D532
- Sunyaev, S. *et al.* (2001) Prediction of deleterious human alleles. *Hum. Mol. Genet.* 10, 591–597
- Crow, J.F. (1999) The odds of losing at genetic roulette. *Nature* 397, 293–294
- Kondrashov, A.S. (1995) Contamination of the genome by very slightly deleterious mutations – why have we not died 100 times over. *J. Theor. Biol.* 175, 583–594
- Keightley, P.D. and Eyre-Walker, A. (2000) Deleterious mutations and the evolution of sex. *Science* 290, 331–333
- Ohta, T. (1992) The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286
- Sousa, A.O. *et al.* (2000) Simulating inbreeding depression through the mutation accumulation theory. *Phys. Act.* 278, 563–570
- Woodworth, L.M. *et al.* (2002) Rapid genetic deterioration in captive populations: causes and conservation implications. *Conserv. Genet.* 3, 277–288
- Laikre, L. (1999) Hereditary defects and conservation genetic management of captive populations. *Zoo Biol.* 18, 81–99
- Rader, R.B. *et al.* (2005) Empirical tests for ecological exchangeability. *Anim. Conserv.* 8, 239–247

- 34 Nielsen, R. (2004) Population genetic analysis of ascertained SNP data. *Hum. Genomics* 1, 218–224
- 35 Wan, Q.H. *et al.* (2004) Which genetic marker for which conservation genetics issue? *Electrophoresis* 25, 2165–2176
- 36 Seddon, J.M. *et al.* (2005) SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Mol. Ecol.* 14, 503–511
- 37 Goossens, B. *et al.* (2006) Genetic signature of anthropogenic population collapse in orang-utans. *PLoS Biol.* 4, 285–291
- 38 Beaumont, M.A. (1999) Detecting population expansion and decline using microsatellites. *For. Genet.* 153, 2013–2029
- 39 Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4, 981–994
- 40 Schlotterer, C. (2003) Hitchhiking mapping – functional genomics from the population genetics perspective. *Trends Genet.* 19, 32–38
- 41 Storz, J.F. *et al.* (2004) Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* 21, 1800–1811
- 42 Vasemagi, A. and Primmer, C.R. (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Mol. Ecol.* 14, 3623–3642
- 43 Llopart, A. and Aguade, M. (1999) Synonymous rates at the RpII215 gene of *Drosophila*: variation among species and across the coding region. *Genetics* 152, 269–280
- 44 Johnson, K.P. and Seger, J. (2001) Elevated rates of nonsynonymous substitution in island birds. *Mol. Biol. Evol.* 18, 874–881
- 45 Keightley, P.D. and Gaffney, D.J. (2003) Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13402–13406
- 46 Hedrick, P. (2005) ‘Genetic restoration’: a more comprehensive perspective than ‘genetic rescue’. *Trends Ecol. Evol.* 20, 109
- 47 Fay, J.C. *et al.* (2001) Positive and negative selection on the human genome. *Genetics* 158, 1227–1234
- 48 Sunyaev, S. *et al.* (2000) Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet.* 16, 198–200
- 49 Lenffer, J. *et al.* (2006) OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. *Nucleic Acids Res.* 34, D599–D601
- 50 McGreevy, P.D. *et al.* (2005) Listing of inherited disorders in animals (LIDA): an online relational database, using non-technical descriptions written by veterinary students. *J. Vet. Med. Educ.* 32, 551–554
- 51 Kondrashov, F.A. *et al.* (2004) Bioinformatical assay of human gene morbidity. *Nucleic Acids Res.* 32, 1731–1737
- 52 Boccia, A. *et al.* (2005) DG-CST (Disease Gene Conserved Sequence Tags), a database of human-mouse conserved elements associated to disease genes. *Nucleic Acids Res.* 33, D505–D510
- 53 Roth, C. *et al.* (2005) The Adaptive Evolution Database (TAED): a phylogeny based tool for comparative genomics. *Nucleic Acids Res.* 33, D495–D497
- 54 Dorus, S. *et al.* (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat. Genet.* 36, 1326–1329
- 55 Akey, J.M. *et al.* (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814
- 56 Kohn, M.H. *et al.* (2000) Natural selection mapping of the warfarin-resistance gene. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7911–7915
- 57 Harr, B. *et al.* (2002) Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12949–12954
- 58 Bonin, A. *et al.* (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* 23, 773–783
- 59 Storz, J.F. and Dubach, J.M. (2004) Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*. *Evolution* 58, 1342–1352
- 60 Campbell, D. and Bernatchez, L. (2004) Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Mol. Biol. Evol.* 21, 945–956
- 61 Gonzalez-Martinez, S.C. *et al.* (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172, 1915–1926
- 62 Wilding, C.S. *et al.* (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* 14, 611–619
- 63 Gibson, G. (2002) Microarrays in ecology and evolution: a preview. *Mol. Ecol.* 11, 17–24
- 64 Whitehead, A. and Crawford, D.C. (2006) Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5425–5430
- 65 Gibson, G. and Mackay, T.F.C. (2002) Enabling population and quantitative genomics. *Genet. Res.* 80, 1–6
- 66 Schlotterer, C. (2004) The evolution of molecular markers - just a matter of fashion? *Nat. Rev. Genet.* 5, 63–69
- 67 Morin, P.A. *et al.* (2004) SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208–216
- 68 Aguilar, A. *et al.* (2004) High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3490–3494
- 69 Waples, R.S. and Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15, 1365–1419
- 70 Kohn, M.H. and Wayne, R.K. (1997) Facts from feces revisited. *Trends Ecol. Evol.* 12, 223–227
- 71 Sunyaev, S. *et al.* (1999) Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. *J. Mol. Med.* 77, 754–760
- 72 Akashi, H. and Eyre-Walker, A. (1998) Translational selection and molecular evolution. *Curr. Opin. Genet. Dev.* 8, 688–693
- 73 Matys, V. *et al.* (2003) TRANSFAC (R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* 31, 374–378
- 74 Przeworski, M. (2002) The signature of positive selection at randomly chosen loci. *Genetics* 160, 1179–1189
- 75 Glinka, S. *et al.* (2003) Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* 165, 1269–1278
- 76 Kauer, M.O. *et al.* (2003) A microsatellite variability screen for positive selection associated with the ‘Out of Africa’ habitat expansion of *Drosophila melanogaster*. *Genetics* 165, 1137–1148
- 77 Thornton, K. and Andolfatto, P. (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* 172, 1607–1619
- 78 Kim, Y. and Nielsen, R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167, 1513–1524
- 79 Palaisa, K. *et al.* (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9885–9890
- 80 Wayne, R.K. *et al.* (1997) Molecular systematics of the Canidae. *Syst. Biol.* 46, 622–653
- 81 Hanson, B. and Weterberg, L. (2002) On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* 11, 2467–2474