MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 22–40 doi: 10.1111/mec.12094

INVITED REVIEWS AND META-ANALYSES

Inferring recent historic abundance from current genetic diversity

PER J. PALSBØLL,*† M. ZACHARIAH PEERY,‡ MORTEN T. OLSEN,§ STEVEN R. BEISSINGER¶ and MARTINE BÉRUBÉ*†

*Marine Evolution and Conservation, Centre of Evolutionary and Ecological Studies, University of Groningen, PO Box 11103 CC, Groningen, The Netherlands, †Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA, ‡Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53703, USA, §Department of Genetics, Microbiology and Toxicology, Stockholm University, SE-106 91, Stockholm, Sweden, ¶Department of Environmental Science, Policy and Management, University of California Berkeley, 130 Mulford Hall, Berkeley, CA 94720-3110, USA

Abstract

Recent historic abundance is an elusive parameter of great importance for conserving endangered species and understanding the pre-anthropogenic state of the biosphere. The number of studies that have used population genetic theory to estimate recent historic abundance from contemporary levels of genetic diversity has grown rapidly over the last two decades. Such assessments often yield unexpectedly large estimates of historic abundance. We review the underlying theory and common practices of estimating recent historic abundance from contemporary genetic diversity, and critically evaluate the potential issues at various estimation steps. A general issue of mismatched spatio-temporal scales between the estimation itself and the objective of the estimation emerged from our assessment; genetic diversity–based estimates of recent historic abundance represent long-term averages, whereas the objective typically is an estimate of recent abundance for a specific population. Currently, the most promising approach to estimate the difference between recent historic and contemporary abundance requires that genetic data be collected from samples of similar spatial and temporal duration. Novel genome-enabled inference methods may be able to utilize additional information of dense genome-wide distributions of markers, such as of identity-by-descent tracts, to infer recent historic abundance from contemporary samples only.

Keywords: coalescent, conservation, mutation–migration–drift equilibrium, population genetic inference

Received 7 April 2012; revision received 1 September 2012; accepted 17 September 2012

Introduction

Abundance, the number of individuals in a population, is a fundamental parameter in evolution, ecology and conservation. Although there are a variety of methods to estimate current abundance (e.g. Lebreton *et al.* 1992; Schwarz & Seber 1999; Lukacs & Burnham 2005; Thomas *et al.* 2010), options to estimate recent historic abundance are far more restricted, limiting assessments of long-term trends in abundance (e.g. Koschinski 2002; Smith & Reeves 2003; O'Connell *et al.* 2007; Clucas 2011; Lund *et al.* 2011). When a population undergoes demographic change, the reduction in genetic diversity is typically subject to a time delay because the rates of processes that alter genetic diversity (i.e. random genetic drift, mutation and selection) change more slowly than demography. This 'time lag' is apparent in the current global human population, which harbours much lower levels of genetic diversity than expected from the current abundance, reflecting its recent expansion from a much smaller ancestral population

Correspondence: Per J. Palsbøll, Fax: +31 (0)50 363 9620; E-mail: p.j.palsboll@rug.nl

(e.g. Cann *et al.* 1987; Horai *et al.* 1992; Rogers & Jorde 1995; Park 2011). In other words, genetic diversity of the modern human population is more representative of the historical, rather than the current, level of abundance.

The observation that current levels of genetic diversity are, in part, a product of historic abundance has been used in studies of natural populations to infer recent historic abundance from the current level of genetic diversity. 'Genetic diversity–based' estimation of recent historic abundance has been particularly popular in studies of endangered species where the current population status and future conservation goals often are defined with reference to the historic abundance prior to anthropogenic effects (e.g. Lotze *et al.* 2011). Indeed, gaining insights into historic abundance is fundamental to elucidating the natural state and dynamics of the biosphere and the ecological processes governing changes therein (e.g. Estes *et al.* 1998; Berger *et al.* 2001; Croll *et al.* 2005; Myers *et al.* 2007; Nicol *et al.* 2010).

Here, we review the population genetic theory underlying genetic diversity–based estimation of recent local historic abundance and the assumptions upon which this general approach relies. We conclude that only under very restrictive conditions is it possible to estimate the recent historic abundance for a local population at a specific narrow period from contemporary genetic diversity with current analytical approaches. The main issue lies with the discrepancy between the temporal (and consequently also the spatial) scale of the estimation and its objective. For the majority of cases, but not all (e.g. Gemmell *et al.* 2004), the objective is an estimate of very recent historic abundance, typically no more than 10–15 generations ago, for a specific population. In contrast, typical estimation procedures yield an estimate of abundance that represents a long-term mean rather than the desired estimate of abundance for the targeted population at a specific time point. This discrepancy between the spatio-temporal scale of the objective and the employed estimation procedure is not unique to genetic diversity–based estimates of recent historic abundance, but occurs in many applications of population genetic theory to questions in ecology and conservation. In addition to this basic issue, assumptions made when deriving recent abundance from contemporary genetic diversity are unlikely to be met in many cases.

Examples of genetic diversity–based estimations of recent historic abundance

During the last two decades, a number of studies have employed estimates of genetic diversity in contemporary populations to infer recent historic abundance of endangered species. Here, we review three high-profile examples that applied this approach to whales (Roman & Palumbi 2003), wolves (Leonard *et al.* 2005) and extinct moas (Gemmell *et al.* 2004) to illustrate the general approach and the inferences drawn from the results.

The large whales have long served as an example of anthropogenic endangerment due to ill-managed exploitation. Abundance of most populations of large whales across the globe was reduced significantly by whaling during the 19th and 20th century. However, the degree of the reduction in abundance caused by whaling remains elusive, as pre-whaling abundances are essentially unknown. In 2003, Roman and Palumbi used previously published mitochondrial control region DNA sequences (published by Andersen *et al.* 2003; Baker *et al.* 1993; Palsbøll *et al.* 1995; Bérubé *et al.* 1998) to infer the historic abundance of humpback whales (*Megaptera novaeangliae*), fin whales (*Balaenoptera physalus*) and minke whales (*Balaenoptera acutorostrata*) in the North Atlantic from current levels of genetic diversity (Roman & Palumbi 2003). Their genetic diversity–based estimates of recent historic abundance for all three species were substantially higher than the estimates of current abundance, in one case by as much as ~24 times (humpback whale). Roman & Palumbi (2003) inferred their genetic diversity–based abundance estimates to represent recent historic (and not current) abundances of these three species, that is, the abundance just prior to the onset of whaling. The genetic diversity–based estimate of pre-whaling abundance in the humpback whale was also substantially higher than other pre-whaling estimates based upon nongenetic data (e.g. backcasting from models using estimated catches, demographic parameters and current abundance estimates). The upper bound from these nongenetic estimates of pre-whaling abundance in the North Atlantic was 25 800 humpback whales (Punt *et al.* 2006), which was ~10 times lower than the genetic diversity–based estimate at 240 000 humpback whales reached by Roman & Palumbi (2003). Implications of the genetic diversity–based estimates of pre-whaling abundance are profound. If the recent historic abundances in the North Atlantic of 240 000 humpback whales, 360 000 fin whales and 265 000 minke whales (Roman & Palumbi 2003) are correct, they suggest that large whales in the North Atlantic have recovered to a much lower degree than currently assumed. High genetic diversity–based estimates of abundance, inferred as representing recent historic abundance, have also been reported for minke whales in the Southern Ocean (Ruegg *et al.* 2010), eastern North Pacific Ocean grey whales, *Eschrichtius robustus* (Alter *et al.* 2007), and indirectly in Southern Ocean blue whales (*Balaenoptera musculus intermedia*) (Sremba

et al. 2012). A larger historic biomass of whales also implies a much greater historic biomass of the prey base, suggesting a significant baseline shift in ocean productivity during the last century (Jackson 2001).

Leonard *et al.* (2005) reached similar conclusions from their genetic diversity–based analysis of historic abundance for the North American grey wolf, *Canis lupus*. The grey wolf is a textbook example of a terrestrial carnivore subjected to intense eradication programs during the last century, resulting in a greatly reduced abundance but to an unknown extent. Leonard *et al.* (2005) sequenced the mitochondrial control region in grey wolf samples from 90- to 150-year-old specimens assumed to represent the pre-exploitation population in the contiguous U.S. Abundance from these historic samples was estimated at 380 000 wolves, much higher than the contemporary abundance in North America, which has been estimated at 60 000 to 70 000 individuals. As was the case for the large whales, Leonard *et al.* (2005) suggested that the conservation goals for U.S. grey wolves be adjusted upward to reflect their genetic diversity–based estimates of recent historic abundance. The ecological ramifications of the findings by Leonard *et al.* (2005) are similar to those of Roman & Palumbi (2003); more wolves imply more abundant prey and suitable habitat in the recent historic past.

Our last example of genetic diversity–based estimates of recent historic abundance is for the moa, *Dinornis* spp., an extinct flightless ratite bird that was endemic to New Zealand. Nongenetic analyses have suggested that abundance prior to human settlement (1000 to 6000 years BP) was at 159 000 individuals (Gemmell *et al.* 2004). Based upon previously published DNA sequences (Huynen *et al.* 2003), Gemmell *et al.* (2004) arrived at genetic diversity–based estimates of 300 000 to 1.4 million individuals between 1000 and 6000 years BP and 3–12 million individuals >6000 years BP. Gemmell *et al.* (2004) interpreted their findings to suggest that the demise of *Dinornis* spp. pre-dates human settlement and the Moa's drastic pre-human settlement decline probably was caused by diseases and epizootics introduced to the islands by immigrating birds.

These three examples, which are typical for the majority of genetic diversity–based estimates of abundance, illustrate two common features: (i) genetic diversity–based estimates of recent historic abundance are generally much larger than estimates of both the current abundance and the presumed historic abundance and (ii) genetic diversity–based estimates are usually interpreted as representing abundance at (or until) a specific point in time and often for a specific population, as well. Some genetic diversity–based estimates of recent historic abundance have since been cited as the *de facto* historic abundance in policy-oriented literature discussing current conservation and management policies (e.g. Myers & Worm 2005; Bolster 2006; Pinnegar & Engelhard 2008; Lotze & Worm 2009; Bonebrake *et al.* 2010; Pershing *et al.* 2010).

Estimating recent historic abundance from contemporary levels of genetic diversity

The popularity of the genetic diversity–based approach is due to the simple and straightforward relationship between the number of mutations (denoted θ) expected for any two random gene copies as the product of the per-generation mutation rate (μ) and the effective population size of the targeted genome (N_e) in a single idealized population. Population genetic theory predicts the following expectation,

$$\theta = xN_e\mu \quad (1)$$

where the value of x depends upon the ploidy and mode of inheritance for the locus under investigation (Ewens 1972; Watterson 1984; Hudson 1990). For a diploid, autosomal and Mendelian-inherited locus x equals 4 and N_e, the effective population size of both sexes. In the case of a haploid, uniparentally inherited locus (such as the maternally transmitted vertebrate mitochondrial genome) x equals 2. In this latter case, N_e denotes the effective population size of the relevant sex (e.g. females in the case of the vertebrate mitochondrial genome). Consequently, if the value of μ is known, then it is, in principle, possible to infer the effective population size from an estimate of θ when the assumptions underlying the above relationship between θ, N_e and μ are upheld. The final estimate of abundance (i.e. census population size, N_c) is inferred from the estimate of N_e by correction for the difference between N_c and N_e. Putting it all together, we obtain the following relationship,

$$\text{Historic } N_c = \left[\frac{N_c}{N_e}\right]\left[\frac{\theta}{x\mu}\right] \quad (2)$$

The rationale for treating N_c as an estimate of recent historic (and not current) abundance (even though N_e is derived from the current observed genetic diversity) is due to the observation that a recent reduction in abundance has little, if any, effect on some measures of genetic diversity such as nucleotide diversity (π, Tajima 1989), heterozygosity (Cornuet & Luikart 1996) and the range of allele sizes [for simple tandem repeated (STR), loci, Garza & Williamson 2001]. Figure 1 depicts the rate of decline in θ (estimated as π) following an instantaneous 95% reduction in the effective size of a population when $\theta = 6.0$. At 200 and 2000 generations since the reduction to 5% of the original population size, the mean estimate of θ is only reduced by 7% and 50%,

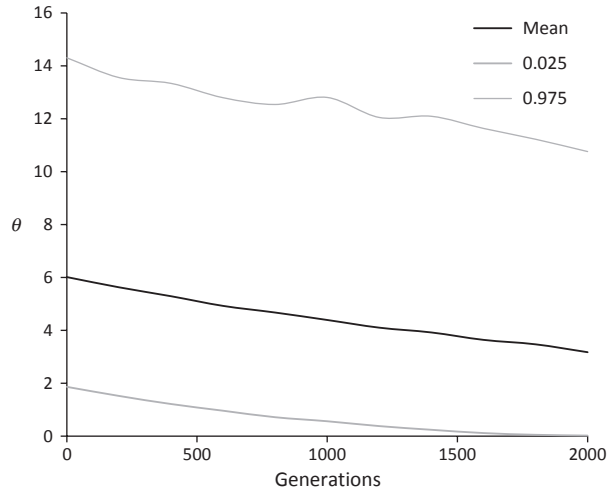

Fig. 1 Rate of decrease in genetic diversity (θ) during the first 2000 generations after a 95% population bottleneck. Data were simulated using coalescent-based simulation software ms (Hudson 2002) emulating the estimates for the humpback whale in Roman & Palumbi (2003) using the following parameters: pre-bottleneck θ at 6, which corresponds to a nucleotide diversity at 0.02 for DNA sequences each of 300 nucleotides and a sample of 200 gene copies. On the x-axis, x denotes the number of generations backward in time when the population was increased instantaneously by a factor of 20 (i.e. from the current value of θ at 0.3 to 6). A total of 10 000 replicates were simulated for each value of x. The 95% confidence interval was estimated as the 97.5 and 0.025 percentile estimates of π (which is equal to θ under an infinite site mutation model).

respectively. The estimate of θ is thus much higher than the expected value of 0.3 for the current abundance. Most genetic diversity–based estimations of recent historic abundance are concerned with much more recent timescales (i.e. <20 generations).

In Table 1, we have outlined the overall estimation procedure and each step necessary to go from an estimate of contemporary genetic diversity (i.e. θ) to an estimate of historic N_c using eqn 2. There are three main components of eqn 2: θ, μ and finally the ratio of N_e to N_c, each of which is estimated separately, using either population genetic, phylogenetic, fossil or contemporary demographic data as well as combinations thereof. In the following section, we briefly outline how each of these estimations is conducted. We then provide a more detailed discussion of some of the potential issues in each of these estimations. Finally, we conclude with a more general discussion of the applicability of the final 'historic' abundance estimate as well as alternate means to obtain more reliable estimates of recent historic abundance from genetic data.

Inferring N_e from an estimate of θ

An estimate of θ can be obtained analytically in cases of a single isolated, unstructured population assuming a simple mutation model (Fig. 2, models A.1), such as the infinite site mutation model (i.e. no nucleotide position is subject to recurrent mutations, Kimura 1969). In this

Table 1 An overview of the estimations and analyses required to derive recent historic abundance from contemporary levels of genetic diversity (i.e. eqn 2 in the main text)

Parameter in eqn 2	Temporal scale of estimation	Data	Analysis	Estimate of
μ				
	Interspecific	Interspecific phylogenetic data	e.g. MODELTEST (Posada & Crandall 1998)	Mutation model parameters
	Interspecific	Interspecific fossil dates and phylogenetic data	e.g. BEAST (Drummond & Rambaut 2007)	Annual mutation rate
	Contemporary	Survival and reproductive rates		Generation time
θ				
	Intraspecific	Intraspecific population genetic data	e.g. MODELTEST (Posada & Crandall 1998)	Mutation model parameters
	Intraspecific	Intraspecific population genetic data	e.g. LAMARC (Kuhner *et al.* 2004), MIGRATE (Beerli & Felsenstein 2001), IMa2 (Hey 2010b), BATWING (Wilson *et al.* 2003)	θ (estimation requires specification of mutation model)
N_e/N_c				
	Contemporary	Sex-specific reproductive and mortality rates	e.g. MARK (White & Burnham 1999)	Contemporary N_e
	Contemporary	Population abundance data	e.g. MARK (White & Burnham 1999), DISTANCE (Thomas *et al.* 2010)	Contemporary N_c

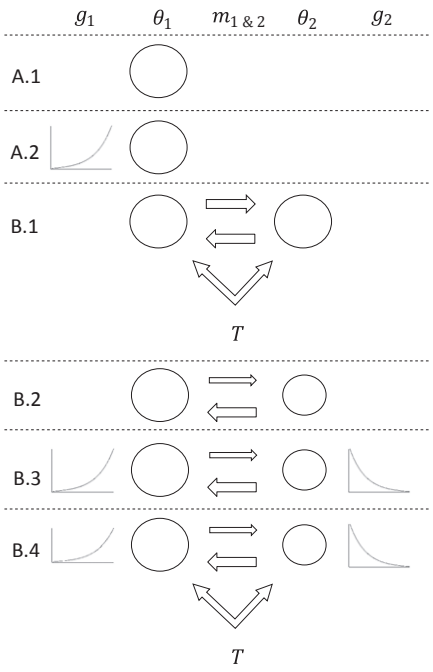

Fig. 2 Population models underlying the most common approaches to infer local genetic diversity. Absence of a symbol implies that the parameter is not specified in the model. Equal-sized population circles (under the heading θ) or migration rate arrows (under the heading m) denote the assumption of identical population sizes and/or migration rates among populations. Different symbol sizes denote that a rate may differ among and in populations. Although an exponential growth rate (g) is depicted, the growth rates may be (for some implementations) linear or instantaneous. Different curves for g in the same model imply that g is estimated separately for each population. T denotes population divergence time. Examples of approaches for the different population models are as follows: Model A.1: Genetic diversity summary statistics, such as π (Nei & Li 1979), number of segregating sites (Tajima 1989) and the variance in number of repeats at STR loci (Di Rienzo *et al.* 1994). Model A.2: Single population-likelihood methods, such as those implemented in BEAST (Drummond & Rambaut 2007), MSVAR (Beaumont 1999), FLUCTUATE (Kuhner *et al.* 1998) and τ (Rogers & Harpending 1992). Model B.1: Likelihood or approximate Bayesian computation (ABC) methods with only two populations with symmetric gene flow and population sizes, such as MDIV (Nielsen & Wakeley 2001). Model B.2 Likelihood or ABC methods with two or more populations, such as MIGRATE-N (Beerli & Felsenstein 2001). Model B.3: Likelihood or ABC methods with two or more populations, such as LAMARC (Kuhner *et al.* 2004). Model B.4: Likelihood or ABC methods with two or more populations, such as IMa2 (Hey 2010a,b) and BATWING (in this implementation divergence times are fixed priors and growth rates identical across populations, Wilson *et al.* 2003).

case, the expectation of θ is the nucleotide diversity (π), which denotes the average number of substitutions per base pair among pairs of gene copies in a population sample (Nei & Tajima 1981). Another example is STR loci evolving in a stepwise manner (Moran 1975). The expectation of θ in this case is equal to twice the population variance in repeat size (Kimmel *et al.* 1996). Analytical derivations are lacking for more complex population (see Fig. 2, models A.2 and B) and mutation models, in which case coalescent-based Markov chain Monte Carlo (MCMC)-likelihood or approximate Bayesian computation (ABC) approaches are employed to estimate θ (Beaumont *et al.* 2002; Kuhner 2009).

Estimating μ, the generational mutation rate, which is necessary to infer N_e from an estimate of θ, is a nontrivial matter in most species and even more so in nonmodel species. Although it is possible to detect novel mutations directly (e.g. in genetic data collected from known parent and offspring pairs), sample sizes of such known relations need to be large due to the scarcity of novel mutations. Even for the most rapidly evolving loci, such as STR loci, novel mutations are still a rare event (10^{-4}–10^{-5} per generation, Ellegren 2000). For nucleotide sequences, the mutation rate is only ~0.5–0.8 mutations per nucleotide site per million years (Millar *et al.* 2008; Subramanian *et al.* 2009) at 'fast'-evolving loci, such as the vertebrate mitochondrial control region, and the average mutation rate across the genome is considerably lower (e.g. ~0.0022 mutations per site per million years in mammals, Kumar & Subramanian 2002). Hence, although in principle it is possible to estimate the mutation rate directly from large samples in known pedigrees (e.g. Millar *et al.* 2008), phylogenetic approaches are more commonly used to estimate mutation rates in nonmodel species (e.g. Shields & Wilson 1987). Phylogenetic approaches calibrate the estimated number of mutations inferred from the observed number of substitutions and implicit or explicit assumptions regarding the mode of mutation against the interspecific divergence times derived from other data (e.g. from dated fossils). The mutation model is usually estimated from the data as well, most commonly selecting among a large number of predefined, nested mutation models using likelihood ratios, Akaike Information Criterion (AIC) score or similar criteria (e.g. MODELTEST by Posada & Crandall 1998). The final result is an annual mutation rate, which then must be converted into a generational mutation rate in order to solve eqn 1.

Generation time is a straightforward entity in species with discrete generations, but can be problematic when generations overlap. Calculating generation time ideally requires age-specific survival and reproductive rates (Deevey 1947; Caswell 2001) for which data often are lacking in nonmodel species. Instead, a population mean age is typically used (e.g. Roman & Palumbi 2003; Alter *et al.* 2007; Ruegg *et al.* 2010).

Deriving N_c from an estimate of N_e

The estimate of 'abundance' derived from eqn 1 is the effective population size (i.e. N_e). The effective population size in population genetics is defined as the size of an 'ideal' population with the observed rate of genetic drift (Wright 1931, 1969) and does thus not represent census population size, the parameter of interest in terms of historic and contemporary abundance. In ecology, the effective population size is usually thought of as the number of breeding individuals that successfully transmit their genes to the next generation (Frankham 1995) and thus should equal the 'genetic' effective population size. In practice, however, the exact relationship is rarely known. Consequently, most studies apply a generic ratio representing a range of estimates (often between 1:4–1:10) obtained from a variety of species (e.g. Frankham 1995; Roman & Palumbi 2003).

Several studies that have inferred recent historic abundance from contemporary genetic diversity in vertebrates were based solely or partly upon DNA sequences from the mitochondrial genome (Hoelzel *et al.* 1993; Kokko *et al.* 1999; Roman & Palumbi 2003; Gemmell *et al.* 2004; Leonard *et al.* 2005; Alter *et al.* 2007). Because the mitochondrial genome in many species is haploid and maternally inherited, a correction of N_c to N_e for sex ratio (and ploidy) is required as well. The sex ratio and ploidy is usually well described in contemporary populations.

The main issues

Estimating θ: Temporal issues

Natural populations are subjected to changing ecological and environmental conditions (e.g. Willerslev *et al.* 1999, 2003; Shapiro *et al.* 2004; Campos *et al.* 2010a,b; Lorenzen *et al.* 2011). As a result, abundance, migration rates and selective agents change over time and space within and among natural populations. The highly variable dynamics experienced by natural populations contrast with the constancy of the *in silico* populations upon which genetic diversity–based estimators of recent historic abundance are based (see Fig. 2). These highly simplified assumptions are necessary to make the estimation of θ tractable and computable. Accordingly, most *in silico* populations are not subject to changes in rates in time or space but instead have constant growth (g), migration (m) and mutation rates. The consequence of this difference between the *in vivo* and *in silico* populations is that the single and final estimate of θ (as well as growth and migration rates, depending on the estimation procedure) represents the most likely, weighted-mean value of θ (and, when applicable, growth and migration rates) over the tempo-spatial range covered by the estimation procedure. In other words, the estimate of θ is *not* necessarily an estimate of θ that applies to a specific point in time and place. We deliberately use the rather vague, term 'weighted mean' because it is not clear what kind of mean the final estimate of θ (and N_e) represents. In a single isolated population, current N_e inferred from genetic diversity is the harmonic N_e (and thus heavily influenced by past bottlenecks). However, in the case of populations subject to both fluctuating demographic and genetic rates, such as migration and mutation rates, this may not necessarily be the case, which is why we use a 'weighted mean' instead of the median.

The time frame of an estimate of θ is the time that has elapsed since the most recent common ancestor, which is $2N_e$ and $4N_e$ generations for haploid and diploid loci, respectively, where N_e differs depending on the mode of inheritance. The time to the most recent common ancestor is thus considerable in many cases. One example is the aforementioned estimate of pre-whaling abundance in North Atlantic whales by Roman & Palumbi (2003). Their estimate of θ (and μ) suggested an effective female population size in North Atlantic humpback whales at ~34 000. Assuming a generation time of 24 years (Barlow & Clapham 1997), the expected time to the most recent common ancestor is 1.6 Myr. Roman & Palumbi (2003) used the coalescent-based MCMC-likelihood method implemented in MIGRATE (Beerli & Felsenstein 2001) to estimate θ, which assume constant migration rates among populations and no population growth (i.e. constant population sizes). Values for mN_e (i.e. the number if migrants per generation) and θ estimated in this manner thus represent the mean values for which the observed data are most likely over a period of approximately 1.6 Myr.

An illustrative example is an assessment of bias for the estimation of mN_e and θ due to recent population divergence conducted by Beerli (2010). The MCMC coalescent-based estimation procedure implemented in MIGRATE-n ignores population divergence time, which essentially equals a change of mN_e from ∞ to a lower value at some point in the past. Using simulated data with θ at 0.0025 and varying population divergence times and migration rates, Beerli's assessment revealed that the effect of ignoring population divergence time prior to $N_e/2$ generations ago on θ was negligible, whereas this was not the case for more recent population divergence times. Using the estimates in Roman & Palumbi (2003) as an illustrative example, Beerli's preliminary assessment translates into an 'effect' time at 408 000 years. The very long time frame involved in this and similar estimations implies that, although it is, in principle, possible to model and thus evaluate most biologically feasible deviations of assumptions causing bias, our knowledge of the relevant

population dynamics during such long time frames is very limited and insufficient to do so.

Estimating θ: Spatial issues

Migrants remove or add genetic diversity to or from local populations, which must be accounted for when the objective is an estimate of local (as opposed to global) abundance from estimates of genetic diversity. Hudson (1990) demonstrated using coalescent-based simulations that the mean estimate of θ in a sample of gene copies collected in a *single* population is equal to the sum of θ for *all* populations combined. This result applies for the case of high ($4mN_e = 1000$) and low ($4mN_e = 0.2$) migration rate (see Fig. 3). In other words, when the level of gene flow is high, all populations are essentially part of the same global population and this is reflected in the level of genetic diversity in each population. In contrast, when mN_e is low, immigrant gene copies are highly divergent from the resident gene copies in the receiving population due to the combined action of mutation and genetic drift. As a result, each immigration event will result in a substantial increase in local genetic diversity (from which N_e is inferred) that does not correspond to the relative demographic increase in local abundance due to immigrant individuals. Consequently, when migration is likely to occur, even at low values of mN_e, an estimation of local N_e from θ must account for migration in order for the estimate to reflect local rather than global abundance. The coalescent-based MCMC approaches implemented in MIGRATE-n (Beerli & Felsenstein 2001), BATWING (Wilson *et al.* 2003), LAMARC (Kuhner *et al.* 2004) or IMA2 (Hey & Nielsen 2004) all yield joint estimates of θ and migration rates in a full matrix of populations, which may also be modelled in an ABC framework.

Relevant, but nonsampled populations are known as ghost populations (Beerli 2004; Slatkin 2005). The most extreme example of ghost populations would be extinct populations, which in most cases will be unknown and unavailable for sampling. The example (from simulated data) depicted in Fig. 4 illustrates that the effects on local genetic diversity due to past migration to and from extinct ghost populations may linger for thousands of years. Even for the species that we probably know best, *Homo sapiens*, new studies constantly change our recent migration history, such as revising the initial immigration of humans into East Asia from 25 000–38 000 BP to 50 000 BP (Rasmussen *et al.* 2011). Our knowledge of the past in natural populations of nonmodel species is far poorer. Ghost populations may be an issue even for large, relatively easy-to-observe species, such as baleen whales, which are and have been subject of intense human attention. Several recent observations in the baleen whales discussed below illustrate that the presence of such ghost populations is a real phenomenon.

The recent historic abundance of eastern North Pacific grey whales was inferred from contemporary estimates of genetic diversity at three to five times the current

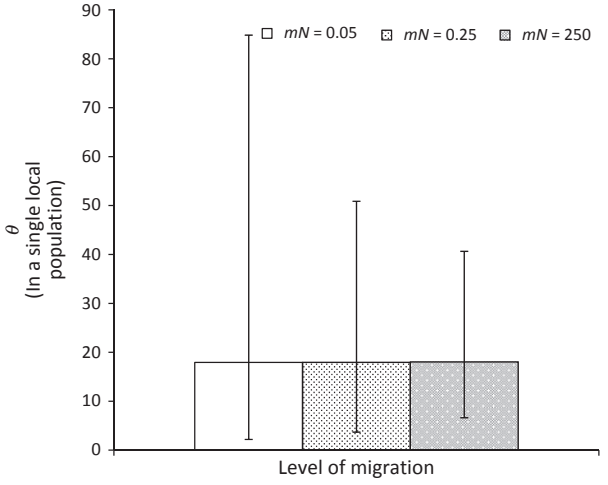

Fig. 3 The effect of migration rates on local levels of genetic diversity (π). Simulations were conducted to emulate the findings presented in Hudson (1990). Data were generated using coalescent-based simulation software ms (Hudson 2002). Sequence data were generated under an Island model with three populations (each with θ = 6) connected by different levels of gene flow. A total of 10 000 data sets were generated for each migration value. Bars denote the 95% confidence interval of the estimate of θ estimated from a sample of 200 gene copies collected from the target population.

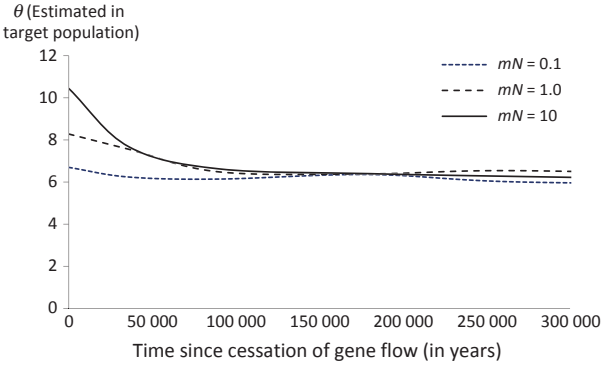

Fig. 4 The lingering effect on local genetic diversity of past connectivity. Data were generated using coalescent-based simulation software ms (Hudson 2002). Data were simulated for two populations each at $\theta = 6$ connected by a migration rate of 0.1, 1.0 or 10, respectively, prior to generation zero, after which gene flow was zero. For each estimate, 100 data sets were generated under an infinite sites mutation model, and the value of θ for the single extant population was estimated as π (Nei & Tajima 1981).

abundance estimated to be ~20 000 individuals (Alter *et al.* 2007). Even though the grey whale type specimen is from the North Atlantic, where several grey whale fossils have been located, grey whales were believed to have gone extinct in the North Atlantic during the 1600s (Reeves & Mitchell 1988). Alter *et al.* (2007) dismissed any effects of gene flow into the North Pacific from the North Atlantic because migration between the North Atlantic and Pacific Ocean was deemed infeasible during the last 114 000 years due to the density of sea ice in the Arctic Ocean. However, during the summer of 2010, a live grey whale was observed in the Mediterranean Sea (Gazo & Chicote 2010). The population origin of the sighted grey whale is unresolved, but the sighting suggests that migration between the North Pacific and North Atlantic possibly supports recent work by Funder *et al.* (2011) and Antoniades *et al.* (2011), suggesting much lower levels of sea ice in the Arctic Ocean 4000–8000 years before present compared to contemporary sea ice levels. A similar observation was made in the common minke whale (*B. bonaerensis*). Ruegg *et al.* (2010) estimated pre-whaling abundance of the common minke whale in the Southern Hemisphere, the only ocean in which this species had been observed, thus greatly simplifying the estimation because population structure could be reasonably ignored (Ruegg *et al.* 2010). In 2011, Glover *et al.* (2010) reported two tissue samples collected from North Atlantic minke whales with mitochondrial DNA sequences matching the common minke whale, one of which appeared to be a first-generation hybrid between the two minke whale subspecies (Glover *et al.* 2010). Phylogeographic analyses in humpback whales and fin whales also suggest occasional trans-equatorial and trans-oceanic migrations (Baker *et al.* 1993; Palsbøll *et al.* 1995; Bérubé *et al.* 1998).

The above examples illustrate clearly the incompleteness in our basic knowledge in aspects that may be critical to inferring recent historic abundance from current levels of genetic diversity, even in large seemingly readily detected and intensively studied organisms.

Issues in estimating μ

An estimate of μ, the generational mutation rate, is required to make inferences of abundance from estimates of current genetic diversity (eqn 1). Studies published to date have relied upon phylogenetic estimates of μ, that is, interspecific calibrations (e.g. Roman & Palumbi 2003; Gemmell *et al.* 2004; Leonard *et al.* 2005; Alter *et al.* 2007; Ruegg *et al.* 2010). This approach is common in taxonomic studies relying on known fossil dates as interspecific time calibration points to get annual mutation rates. In the case of taxonomy, the objective (i.e. interspecific divergence times) and the estimation method (i.e. phylogenetic calibration of mutation rates using known interspecific time points) are both on similar temporal scales (e.g. Hasegawa *et al.* 1985; Ball & Avise 1992; Brunner *et al.* 2001; Thorne & Kishino 2002; Barker *et al.* 2004; Yang & Rannala 2006; Meredith *et al.* 2011). When estimating recent historic abundance, however, there is a substantial difference between the temporal time frame of the estimation (i.e. of μ) and the objective (i.e. recent historic abundance). The possible effect of this difference in time frames is exacerbated by the fact that relatively rapidly evolving loci are utilized in intraspecific studies.

A case in point is the mitochondrial control region, arguably the most common and widespread DNA sequence employed in intraspecific assessments of genetic diversity in nonmodel vertebrates (Moritz 1994; Galtier *et al.* 2009). The mitochondrial control region exhibits highly heterogeneous mutation rates among sites, with so-called hyper-variable regions dominating in the 5′ end (Aquadro & Greenberg 1983; Hoelzel 1993; Ho *et al.* 2011). A general discrepancy has emerged when comparing the mutation rates obtained in the traditional manner (i.e. the phylogenetic approach) to the rates estimated from sequences collected in pedigrees of closely related individuals (Howell *et al.* 2003a,b, 2005; Irwin *et al.* 2009; Loogvali *et al.* 2009; Phillips *et al.* 2009; Goto *et al.* 2011; Klutsch *et al.* 2011) as well as temporally spaced samples, in so-called ancient DNA analyses (Ho *et al.* 2007, 2008). In general, the mutation rates derived from pedigrees or ancient DNA studies have been ~10 times higher than the phylogenetic-derived rates (e.g. Millar *et al.* 2008). Possible explanations for this discrepancy are several and the subject of considerable current debate (Ho & Larson 2006; Emerson 2007; Navascues & Emerson 2009; Ho & Lanfear 2010; Ho *et al.* 2011). While the phylogenetic-derived mutation rate may be underestimated due to recurrent mutations (Santos *et al.* 2005; Phillips *et al.* 2009), the pedigree-derived mutation rate may be overestimated due to transient mutations that are lost almost immediately (Irwin *et al.* 2009; Klutsch *et al.* 2011). Ho and colleagues have published a series of studies addressing this particular aspect (Ho *et al.* 2005, 2007, 2011; Ho & Larson 2006; Ho & Lanfear 2010), arguing that mutation rates are subject to temporal rate variation due to temporal variation in selection intensity (e.g. Loogvali *et al.* 2009). Other authors have utilized other slower-evolving regions in the mitochondrial genome as a means to conduct an *ad hoc* correction for recurrent substitutions in the mitochondrial control region (Alter & Palumbi 2009; Phillips *et al.* 2009).

Irrespective of the underlying cause, it is becoming increasingly clear that phylogenetic methods tend to underestimate μ by as much as one order of magni-

tude, which consequently implies a corresponding overestimation of recent historic abundance, all other factors being equal (Ho *et al.* 2008, 2011). The above-mentioned *ad hoc* corrections yielded new, increased estimates of μ resulting in a corresponding lower estimate of recent historic abundance. For instance, Alter & Palumbi (2009) estimated μ for the mitochondrial control region at 2.6 times higher than previous estimates, implying a revised genetic diversity–based estimate of recent historic abundance of humpback whales in the North Atlantic from 240 000 to 90 000–100 000 individuals.

The use of temporally spaced samples (i.e. ancient and modern DNA samples) enables the estimation of the mutation rate over a more appropriate time frame (Drummond *et al.* 2002), but assumptions of current approaches (Fig. 2A.2) are very basic (e.g. ignoring gene flow), which may lead to incorrect inferences (Navascues & Emerson 2009).

Natural selection, which potentially affects the observed substitution rate as well as N_e, alters the rate of genetic drift relative to neutral expectations. If the degree of selection is similar across all lineages used in the estimation of μ, then a phylogenetic estimation of μ should, in principle, be unbiased. However, if the timing and strength of selection is unequal among the lineages (e.g. species or populations) employed in the estimation of μ and θ, then the resulting estimate of θ may be biased (Galtier & Boursot 2000; Torroni *et al.* 2001) unless the rate estimate is corrected for fluctuations in the timing and strength of selection (Tamura & Kumar 2002; Soares *et al.* 2009). Although genetic diversity–based estimates of recent historic abundance to date have assumed all loci to be selectively neutral, it is likely that many loci employed are subject to selection —both the presumed neutral mitochondrial control region (Casane *et al.* 1997; Howell *et al.* 2004, 2007; Haney *et al.* 2010) and nuclear loci, which often are introns (e.g. Palumbi & Baker 1994; Bouck & Vision 2007; Jackson *et al.* 2009; Li *et al.* 2010; Gante *et al.* 2011; Kemppainen *et al.* 2011).

Issues in estimating demographic parameters

Estimates of both generation time and the ratio of effective to census population size are needed to estimate recent historic abundance from contemporary genetic diversity. Generation time is needed to translate annual mutation rates into generational mutation rates, whereas the ratio of effective to census population size is needed to convert estimates of N_e (derived from the estimates of θ and μ) into estimates of abundance. Generation time is not a fixed entity but can change over time and space due to the variation in demographic and environmental processes, such as population density and habitat changes. These fluctuations, in turn, alter the age at first reproduction, overall reproductive output and mortality rates (Deevey 1947; Caswell 2001). Hence, a full-fledged estimation of generation time would require a complete Leslie matrix or life table with detailed demographic data on reproductive and mortality rates, which are unavailable in many non-model species. In addition, demographic rates vary over time and across the species' range. For example, both age distribution and age-specific fecundity rates are unknown in most species of baleen whales. Some species appear to have very long life spans of more than one hundred years (Garde *et al.* 2007; Arrigoni *et al.* 2011; George *et al.* 2011). Even in some of the best studied species, such as humpback whales, with shorter life spans (such as 30–50 years), estimates of mean age at maturity differ by a factor of two (12 and 24 years, respectively) depending on the methods used to determine age (Chittleborough 1965; Barlow & Clapham 1997). The difference between these two 'generation times' results in a doubling of μ, which in turn reduces the resulting abundance estimate by 50%.

In order to convert an estimate of N_e to abundance (i.e. census population size or N_c), the ratio of the two population size measures must be known. Census population size is a relatively straightforward entity (but difficult to obtain in many cases), whereas the same is not the case for the effective population size which (in this context) is a population genetic entity. Hence, in order to infer N_c from an estimate of N_e, a rigorous estimate of the ratio of current effective population size is warranted. Indeed, in the absence of temporally-spaced genetic samples, estimating N_e requires values for reproductive variance among individuals, as age of reproduction, and age-specific reproductive and survival rates (Waples 2002).

Is it feasible to infer recent historic abundance from current levels of genetic diversity?

All estimation relies upon a model, in other words, a set of explicit (or implicit), often highly simplified assumptions. Such simplification consists of reducing the number of nuisance parameters (i.e. parameters that are not of interest *per se* but which must be accounted for), such as g, μ and mN_e. These simplifications are necessary to render a problem analytically (or computationally) tractable and to achieve a reasonably precise estimate. Ideally, a model is simplified by reducing the number of parameters to only those that are essential (Haldane 1964). Which parameters to retain depend upon the objective of the estimation and targeted level of precision (i.e. each nuisance parameter fixed at a single value results in an

increased precision). Population genetic inference methods commonly used in ecology and conservation are often effective (compared to many other methods used in evolution, ecology and conservation) because they are based on highly simplified assumptions, but such simplifications can also become an Achilles Heel.

The population genetic inference methods that currently are employed to estimate recent historic abundance from current levels of genetic diversity were not developed with this specific objective in mind. Most of the parameter values necessary to infer recent historic abundance from current genetic diversity using eqns 1 and 2 are usually poorly known and fixed at a single value during the estimation (e.g. the mutation, migration and growth rates and generation time). The consequence is inference methods that yield estimates which can best be described as long-term and spatially broad weighted means. The very long time span often covered in the estimation of θ (and μ) greatly increases the spatial scale of the estimation due to the effect of past and present gene flow upon local levels of genetic diversity, necessitating assumptions (and fixed nuisance parameter values) for a multitude of cellular and population processes far back in time (see Fig. 5).

These observations beg the key question: Does a genetic diversity–based estimate of N represent the abundance during recent historic time? Although there is no single universal answer to this question, we believe the answer in the majority of cases is probably 'no'. Exceptions are those few (if any) cases where the targeted population, those populations to which it is connected by migration, and the employed genetic markers behave as assumed by the estimation procedure and do so across the entire spatial and temporal scale covered by the estimation. This is rarely, if ever, the case for natural populations. Thus, an assessment of the extent of bias due to biologically plausible deviations from underlying assumptions should be undertaken. Indeed, some deviations (or combinations thereof) may have little effect on the final estimate of recent historic abundance, whereas other deviations may substantially bias the final estimate.

The reduction of nuisance parameters also implies that the uncertainty in the final abundance estimate likely is underestimated. Thus, most genetic diversity–based estimates of recent historic abundance appear overly precise.

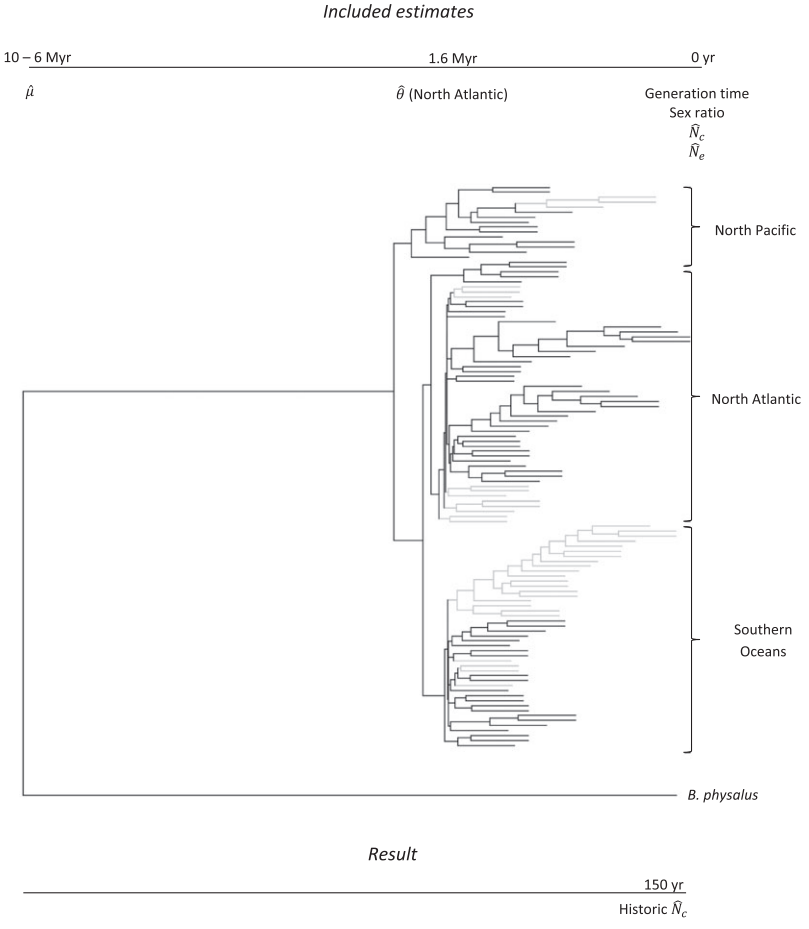

Fig. 5 Illustration of the different temporal and spatial scales involved in a typical estimation of recent historic abundance of a local population. The depicted genealogy is estimated from the mitochondrial DNA sequence data used by Roman & Palumbi (2003) to estimate the recent historic abundance in North Atlantic humpback whales. The data were originally published by Baker *et al.* (1993) and Palsbøll *et al.* (1995). The mtDNA lineages depicted in grey were obtained from whales sampled in a different ocean basin than that which the specific clade was assigned to. These branches were inferred as the results of rare interoceanic gene flow (Baker *et al.* 1993; Palsbøll *et al.* 1995). The genealogy is the majority-rule consensus tree estimated from 1000 bootstrapped samples and an unrooted neighbour-joining tree with a transition/transversion ratio of 20, as implemented in the PHYLIP 3.69 computer package (Felsenstein 1993). The abbreviation yr denotes years and Myr, million years.

In conclusion, only under the most restrictive of circumstances—namely when populations under study behave as assumed by models employed in the estimation—is a genetically derived estimate of recent historic abundance likely to be correct. While increasing the amount of data (loci and/or samples) will produce a more precise 'abundance' estimate, such an estimate is, in most cases, unlikely to represent local abundance at a specific recent time period as all the above-mentioned simplifying assumptions apply across all loci and samples.

Look before leaping to conclusions

The highly simplistic assumptions underlying most population genetic inference methods that are applied to infer recent historic local abundance are unlikely to apply to most natural populations, and hence, one may conclude that estimating recent historic abundance from contemporary genetic diversity is ill-advised. However, before dismissing such results altogether, it is worthwhile to conduct an assessment of biologically relevant deviations from the underlying assumptions and their effect upon the final estimate of recent historic abundance. If the effect of such deviations is small, then the final abundance estimate derived from an assessment of contemporary genetic diversity may be sound and reasonably accurate. Assessing the effects of different parameter values is relatively straightforward for some parameters (e.g. generation time and μ) when there is a relatively simple relationship with N_e. Figure 6 is an example of such an assessment, which illustrates the large impact of changes in μ relative to generation time. Given the extensive body of population genetic simulation software available (e.g. Hudson 2002; Peng & Kimmel 2005; Excoffier & Foll 2011), it has become relatively straightforward to simulate most population histories and models under a variety of mutation models, thereby generating *in silico* genetic data from which to estimate the 'historic' abundance (reviewed in Hoban *et al.* 2012). Such simulations may not only serve to assess the effect of deviations from the assumptions underlying the estimation, but serve equally to plan a study by specifying the amount of data required to achieve a reasonably correct and precise estimate of historic abundance at the targeted tempo-spatial scale.

An estimate of recent historic abundance inferred from contemporary levels of genetic diversity may also provide the means for predicting a range of statistical characteristics for other genetic data, such as complementary STR data (e.g. genetic bottleneck tests or skyline plots). Additional tests will strengthen the conclusion or reveal inconsistencies. In undertaking additional testing, it is crucial to ensure that different tests or estimates are truly independent. This is often not the case for genetic analyses, which in many cases are based upon related data or related aspects of those data, and hence, care should be undertaken in the analysis and interpretation of such additional estimations (e.g. Alter *et al.* 2012).

Returning to our three introductory examples—large whales in the North Atlantic (2003), grey wolves in North America (Leonard *et al.* 2005) and New Zealand

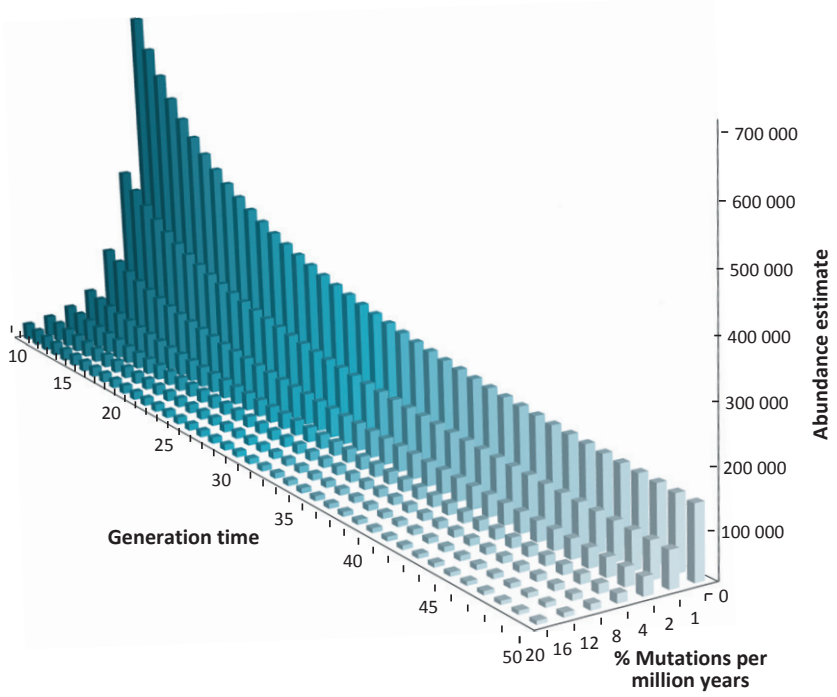

Fig. 6 Relative effect of changes in parameter values of generation time and μ. The graph was modelled after Roman & Palumbi (2003, Fig. 2 on page 509) who estimated the pre-whaling abundance of humpback whales in the North Atlantic from the contemporary level of genetic diversity at the mitochondrial control region. We extended the mutation rate to 20% per million years (the original figure goes up to 7%), because pedigree estimates of mutation rates in the mitochondrial control region suggest that these may be one order of magnitude higher than phylogenetic estimates of mutation rates used in Roman & Palumbi (2003). Estimates of historic abundance from other data sources suggest a pre-whaling abundance at 25 000 individuals (Punt *et al.* 2006), which would be consistent with a value of μ greater than the values used by Roman & Palumbi (2003) but less than that estimated from human pedigrees.

moas (Gemmell *et al.* 2004), we ask whether the estimates presented are representative of recent historic abundance and if so are they likely to be accurate? Our review of present knowledge and the analytical approaches used indicates numerous potential issues for all three cases. A full-fledged, well-supported assessment of the precision and the degree and direction of bias in the estimates of historic abundance for each of these cases is beyond the scope of this review, but some general comments are possible. All three studies were based upon a single locus (mitochondrial DNA) with the added uncertainties single locus estimates entail. The three studies differ in temporal and spatial scales aiming at estimating abundance at a specific place and time. The study in moas, by Gemmell *et al.* (2004), aimed at a broad temporal and spatial scale and is thus likely less susceptible to some of the temporal and spatial issues we have outlined above as the objective was a global long-term estimate. Leonard *et al.*'s (2005) study of grey wolves in North America was also inferred as representing the historic abundance at a 'global scale', and hence, spatial issues are likely less of a concern, although immigration to and from Asia (across the Bering Sea) and Greenland/Scandinavia via the Arctic might occur. However, the estimate was viewed as representing abundance prior to the anthropogenic extermination of grey wolves. For reasons explained above, this is unlikely to be the case. The third study, the abundance of whales in the North Atlantic Ocean at the onset of whaling (Roman & Palumbi 2003), presented the genetic diversity–based estimate of recent historic abundance as representing the local abundance for a specific population (the North Atlantic) and narrow time period (i.e. just prior to the onset of whaling). In this latter case, all the temporal and spatial issues we have outlined in this review apply to this study, in addition to the issues raised about estimates of demographic and mutation rates.

Nevertheless, each of the three studies represents one hypothesis for past abundance, and large discrepancies among different estimates (from genetic as well as non-genetic data) highlight the incompleteness of our current understanding of past abundances. Rigour would be added to the conclusions if estimates derived from different sources and methods converged. Any single estimate of past abundance based only on current genetic diversity should be viewed as tentative.

How then may we estimate recent historic abundance from genetic data?

Significant technical advances in the field of genomics now permit the production of massive amounts of genetic data from natural populations at an ever decreasing effort and cost. Accordingly, limitations in estimating recent historic abundance lie mainly in the analytical approach, in particularly the abovementioned discrepancy between the tempo-spatial scale of the objective and the estimation.

One possible solution is to resolve the discrepancy in the spatial and temporal scales between the objective and the estimation. This may be carried out either (i) by expanding the scope of the objective or inference or (ii) by narrowing the tempo-spatial scale of the estimation to match that of the objective. Expanding the scope of the estimation of abundance to the long-term global average (rather than local) abundance may be sufficient for some objectives (e.g. Gemmell *et al.* 2004; Leonard *et al.* 2005; Ruegg *et al.* 2010), but not others (Roman & Palumbi 2003). When the objective is a long-term and global estimate of abundance, we believe that current genetic approaches can yield reasonably accurate estimates provided that estimates of μ and the necessary demographic parameters are sound.

The temporal scale of the inference may be narrowed by estimating the relative change in θ for each intercoalescent interval in a population sample of orthologous gene copies which generates so-called Bayesian skyline/ride plots (Drummond *et al.* 2005; Drummond & Rambaut 2007; de Bruyn *et al.* 2011; Wu & Drummond 2011). This approach may be conducted with contemporary samples only, but benefits greatly by the inclusion of 'ancient' DNA (aDNA) obtained from carbon-dated samples (e.g. Shapiro *et al.* 2004; Campos *et al.* 2010a,b; Lorenzen *et al.* 2011). The aDNA samples add intraspecific calibration dates (to complement interspecific fossil calibration dates), enabling an estimate of the mutation rate based upon both inter- and intraspecific time calibration points (Ho *et al.* 2008). A key obstacle in terms of inferring recent historic abundance from Bayesian skyline plots is the low information content of the most recent coalescent intervals (Ho & Shapiro 2011). Accordingly, Bayesian skyline plots but will often fail to detect very recent (i.e. 5–15 generations ago) demographic changes (but see Hoffman *et al.* 2011) and the precision of the estimate of θ (and consequently N_e) is low (e.g. Hoffman *et al.* 2011 who estimated the 95% credible interval of pre-exploitation abundance at 0.8–2.3 million individuals). Hence, Bayesian skyline plots are better suited to detect older and prolonged changes in N_e (Fontaine *et al.* 2010; Herman & Searle 2011; Hope *et al.* 2011; Jezkova *et al.* 2011; Liu *et al.* 2011; Morgan *et al.* 2011; Mourier *et al.* 2012). This also appears to be the case with an equivalent approach developed for STR loci, MSVAR (Beaumont 1999; Bourke *et al.* 2010; Girod *et al.* 2011).

However, Bayesian skyline plots (and estimations using MSVAR) are based upon a single population model and thus ignore population structure (and changes in

migration rates), which may bias results substantially (Chikhi *et al.* 2010). Translating changes in genetic diversity estimated from Bayesian skyline plots into census abundance requires the use of the same conversion factors, such as the generation time and the ratio of effective to census population size. Thus, while Bayesian skyline (using aDNA) may deal better with issues surrounding mutation rates, they are subject to many of the same issues affecting more standard genetic diversity–based estimators of abundance. In addition, taking advantage of aDNA samples requires historic samples from the relevant areas and the technical skills and a laboratory set-up to work with degraded DNA (Cooper & Poinar 2000).

Another manner by which to narrow the temporal and spatial scale is to employ samples from around the (historic) time point when an estimate of abundance is required. In such case, an estimate of N_e may be inferred from the degree of linkage disequilibrium (LD) in a random population sample (Hare *et al.* 2011; Waples *et al.* 2011). Individuals in small populations share more recent ancestors and therefore display increased levels of linkage across the genome, which may be employed to estimate the N_e of the parental generation (reviewed in Hare *et al.* 2011). In an ideal population, that is, with discrete generations, it is relatively straightforward to determine the point in time that such an estimate of N_e applies. The point in time is less certain for species with overlapping generations and poorly known demographic structure (e.g. Hare *et al.* 2011), but the window is restricted to a small number of generations. Most current implementations to estimate N_e from LD are based upon a single ideal population model and thus ignore migration, although some methods do attempt to adjust for movements among populations (Vitalis & Couvet 2001). Waples & England (2011) assessed the impact of violation of assumptions fluctuations on the final estimate N_e due to migration and structured samples, when estimators ignore these processes. Not surprisingly, both processes may have substantial effects upon the final estimate of N_e. For instance, generational migration rates above 5–10% (which translates into very low annual migration rates in long-lived species) strongly influence the estimate of N_e (Waples & England 2011). The authors suggest that immigrants may be identified and removed by other means, such as assignment methods (Paetkau *et al.* 1995; Piry *et al.* 2004), although the rigour of individual multilocus assignments decreases with reduced divergence (Waples & Gaggiotti 2006; Hall *et al.* 2009), and consequently, this kind of adjustment may not be infeasible at such high rates of migration (i.e. above 5–10% per generation). The advantage of N_e estimates from LD is the very narrow time scale of the estimation (i.e. the last few generations depending upon the approach and data used), which greatly reduces the uncertainty and possible influence from other processes (e.g. migration and mutation, and variations in those and demographic parameters) that may bias the final abundance estimate. However, in most cases, the 'historic' abundance of interest will be farther back in time than one or two generations, which then implies the need for acquiring 'historic' samples along with all the technical difficulties associated with collecting genetic data from such samples. Last, but not least, the estimation of N_e from LD requires substantial sample sizes in most cases but especially for large undepleted populations (Waples & Do 2008), which may altogether prevent estimating N_e in this manner from a few rare historical samples.

Novel opportunities in the age of genomics and ABC?

Perhaps the most exciting prospects for estimating recent historic abundance arise from novel opportunities offered by massive parallel sequencing technologies as well as new analytical approaches enabling the assessment of more and complex models. These new technologies provide the opportunity to increase the number of loci by several orders of magnitude. As we have pointed out above, adding more data will increase the precision of the final abundance estimate, but for many inference methods the estimate will represent a long-term mean of abundance and not the local abundance at a specific recent time period. However, the greatly increased genome-wide density of data adds new possibilities in terms of estimating the degree of linkage among markers along chromosomes provided that a (partial) genome backbone is available. Such data enable the estimation of the lengths and distribution of linkage tracts in each sample. The smaller the population, the longer the linkage tracts [also known as identity-by-descent (IBD) tracts], because most individuals share a very recent common ancestor, resulting in low recombination rates. The distribution of the lengths of IBD tracts in an individual's genome reflects past and present population processes, such as selection, migration and effective population size. Recently, Pool & Nielsen (2009) showed how changes in migration rates affect the distribution of lengths of 'migrant' tracts. The same should be the case for changes in population size, which will be reflected in the genome-wide distribution of IBD tract lengths (Chapman & Thompson 2003). The degree of LD will be highest for small N_e's, which implies longer IBD tracts and greater precision at low N_e's, whereas the opposite will be the case for large N_es (Chapman & Thompson 2003). Accordingly, the utility of the genome IBD tract-

lengths approach will depend, in part, upon the size of the historic population with low precision at large historic population sizes (Pool *et al.* 2010).

Although these new technological advances offer great potential, the speed and computational resources required to make inferences under sufficiently realistic population and mutation models is an associated and purely practical issue. Even the current (comparatively low coverage) multilocus DNA sequence data set (e.g. Alter *et al.* 2007; Ruegg *et al.* 2010) poses a computational challenge for likelihood-based estimation methods. The adoption of massive parallel sequencing technologies (Davey *et al.* 2011) will lead to much more data per specimen. Increasing the amount of data usually increases the precision and rigour of the estimation. However, likelihood MCMC approaches commonly employed to estimate θ from population genetic data can quickly become computationally infeasible as the amount of data increases (Wegmann *et al.* 2009). Sequential Markov coalescent–based (McVean & Cardin 2005; Excoffier & Foll 2011) or ABC (Beaumont *et al.* 2002; Hamilton *et al.* 2005; Wegmann *et al.* 2010) approaches appear to constitute a promising alternative to the likelihood approaches (Lopes & Boessenkool 2010). The computational efficiency afforded by these non-likelihood approaches will make it feasible to utilize more of the information contained in the large amounts of data generated by these parallel sequencing methodologies, such as IBD tracts and the allele frequency spectrum which may be used to estimate demographic changes (Nielsen 2000; Polanski & Kimmel 2003; Caicedo *et al.* 2007; Gusev *et al.* 2009; Gutenkunst *et al.* 2009; Crisci *et al.* 2012). However, even though non-likelihood-based approaches provide additional flexibility (and thus are able to accommodate more complex models), such estimations need careful consideration and a realistic number of nuisance parameters (Bertorelle *et al.* 2010).

Acknowledgements

PJP was in part supported by a Hrdy Fellowship in Conservation Biology from Harvard University during the writing of the manuscript. SRB was supported by NSF (DEB-1051342 and CNH 1115069). We are indebted to Drs. Oscar Gaggiotti, Allendorf, Luis Pastene and Finn Larsen as well as two anonymous reviewers for helpful comments which greatly improved the manuscript.

References

Alter SE, Palumbi SR (2009) Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome *b* in three species of baleen whales. *Journal of Molecular Evolution*, **68**, 97–111.

Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 15162–15167.

Alter SE, Newsome SD, Palumbi SR (2012) Pre-whaling genetic diversity and population ecology in eastern pacific gray whales: insights from ancient DNA and stable isotopes. *PLoS ONE*, **7**, e35039.

Andersen LW, Born EW, Dietz R, Haug R, Øien N, Bendixen C (2003) Genetic population structure of minke whales *Balaenoptera acutorostrata* from Greenland, the North East Atlantic and the North Sea probably reflects different ecological regions. *Marine Ecology Progress Series*, **247**, 263–280.

Antoniades D, Francus P, Pienitz R, St-Onge G, Vincent WF (2011) Holocene dynamics of the Arctic's largest ice shelf. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 18899–18904.

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*, **103**, 287–312.

Arrigoni M, Manfredi P, Panigada S, Bramanti L, Santangelo G (2011) Life-history tables of the Mediterranean fin whale from stranding data. *Marine Ecology*, **32**, 1–9.

Baker CS, Perry A, Bannister JL *et al.* (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proceedings of the National Academy of Science of the United States of America*, **90**, 8239–8243.

Ball RM, Avise JC (1992) Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *Auk*, **109**, 626–636.

Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J (2004) Phylogeny and diversification of the largest avian radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 11040–11045.

Barlow J, Clapham PJ (1997) A new birth-interval approach to estimating demographic parameters of humpback whales. *Ecology*, **78**, 535–546.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.

Beerli P (2010) Violation of assumptions, or are your migration estimates wrong when the populations split in the recent past? http://popgen.sc.fsu.edu/Migrate/Blog/Entries/2010/8/15_Violation_of_assumptions,_or_are_your_migration_estimates_wrong_when_the_populations_split_in_the_recent_past.html, Florida State University.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Berger J, Stacey PB, Bellis L, Johnson MP (2001) A mammalian predator-prey imbalance: Grizzly bear and wolf extinction affect avian neotropical migrants. *Ecological Applications*, **11**, 947–960.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Bérubé M, Aguilar A, Dendanto D *et al.* (1998) Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus, 1758): analysis of mitochondrial and nuclear loci. *Molecular Ecology*, **7**, 585–600.

Bolster WJ (2006) Opportunities in marine environmental history. *Environmental History*, **11**, 567–597.

Bonebrake TC, Christensen J, Boggs CL, Ehrlich PR (2010) Population decline assessment, historical baselines, and conservation. *Conservation Letters*, **3**, 371–378.

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology*, **16**, 907–924.

Bourke BP, Frantz AC, Lavers CP *et al.* (2010) Genetic signatures of population change in the British golden eagle (*Aquila chrysaetos*). *Conservation Genetics*, **11**, 1837–1846.

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Arctic charr (*Salvelinus alpinus L.*) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573–586.

de Bruyn M, Hoelzel AR, Carvalho GR, Hofreiter M (2011) Faunal histories from Holocene ancient DNA. *Trends in Ecology & Evolution*, **26**, 405–413.

Caicedo AL, Williamson SH, Hernandez RD *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *Plos Genetics*, **3**, 1745–1756.

Campos PF, Kristensen T, Orlando L *et al.* (2010a) Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology*, **19**, 4863–4875.

Campos PF, Willerslev E, Sher A *et al.* (2010b) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5675–5680.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature*, **325**, 32–36.

Casane D, Dennebouy N, deRochambeau H, Mounolou JC, Monnerot M (1997) Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of lagomorphs. *Molecular Biology and Evolution*, **14**, 779–789.

Caswell H (2001) *Matrix Population Models: Construction, Analysis, and Interpretation*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Chapman NH, Thompson EA (2003) A model for the length of tracts of identity by descent in finite random mating populations. *Theoretical Population Biology*, **64**, 141–150.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics*, **186**, 983–995.

Chittleborough RG (1965) Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). *Australian Journal of Marine and Freshwater Research*, **16**, 33–128.

Clucas R (2011) Long-term population trends of Sooty Shearwater (*Puffinus griseus*) revealed by hunt success. *Ecological Applications*, **21**, 1308–1326.

Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. *Science*, **289**, 1139.

Cornuet JM, Luikart G (1996) Description and power of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Crisci JL, Poh Y-P, Bean A, Simkin A, Jensen JD (2012) Recent progress in polymorphism-based population genetic inference. *Journal of Heredity*, **103**, 287–296.

Croll DA, Maron JL, Estes JA, Danner EM, Byrd GV (2005) Introduced predators transform subarctic islands from grassland to tundra. *Science*, **307**, 1959–1961.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Deevey ES (1947) Life tables for natural populations of animals. *Quarterly Review of Biology*, **22**, 283–314.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Science of the United States of America*, **91**, 3166–3170.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**, 1307–1320.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**, 551–558.

Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Systematic Biology*, **56**, 337–345.

Estes JA, Tinker MT, Williams TM, Doak DF (1998) Killer whale predation on sea otters linking oceanic and nearshore ecosystems. *Science*, **282**, 473–476.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **21**, 87–112.

Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, **27**, 1332–1334.

Felsenstein J (1993) *PHYLIP. Manual and Program Available from Joe Felsenstein*. University of Washington, Seattle, Washington.

Fontaine MC, Tolley KA, Michaux JR *et al.* (2010) Genetic and historic evidence for climate-driven population fragmentation in a top cetacean predator: the harbour porpoises in European water. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **277**, 2829–2837.

Frankham R (1995) Effective population size adult population size ratios in wildlife—a review. *Genetical Research*, **66**, 95–107.

Funder S, Goosse H, Jepsen H *et al.* (2011) A 10,000-year record of Arctic Ocean sea-ice variability-view from the Beach. *Science*, **333**, 747–750.

Galtier N, Boursot P (2000) A new method for locating changes in a tree reveals distinct nucleotide polymorphism vs divergence patterns in mouse mitochondrial control region. *Journal of Molecular Evolution*, **50**, 224–231.

Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Gante HF, Alves MJ, Dowling TE (2011) Paralog-specific primers for the amplification of nuclear Loci in Tetraploid Barbels (Barbus: Cypriniformes). *Journal of Heredity*, **102**, 617–621.

Garde E, Heide-Jørgensen MP, Hansen SH, Nachman G, Forchhammer MC (2007) Age-specific growth and remarkable longevity in narwhals (*Monodon monoceros*) from West Greenland as estimated by aspartic acid racemization. *Journal of Mammalogy*, **88**, 49–58.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Gazo M, Chicote CA (2010) *Gray Whale Sighting Confirmed in Barcelona*. MarMAM List server, https://lists.uvic.ca/mailman/listinfo/marmam.

Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**, S430–S432.

George JC, Follmann E, Zeh J, Sousa M, Tarpley R, Suydam R, Horstmann-Dehn L (2011) A new way to estimate the age of bowhead whales (Balaena mysticetus) using ovarian corpora counts. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **89**, 840–852.

Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics*, **188**, 165–179.

Glover KA, Kanda N, Haug T *et al.* (2010) Migration of Antarctic minke whales to the Arctic. *PLoS ONE*, **5**, e15197.

Goto H, Dickins B, Afgan E *et al.* (2011) Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biology*, **12**, R59.

Gusev A, Lowe JK, Stoffel M *et al.* (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, **19**, 318–326.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Haldane JBS (1964) A defense of beanbag genetics. *Perspectives in Biology and Medicine*, **7**, 343–359.

Hall LA, Palsbøll PJ, Beissinger SR *et al.* (2009) Characterizing dispersal patterns in a threatened seabird with limited genetic structure. *Molecular Ecology*, **18**, 5074–5085.

Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417.

Haney RA, Silliman BR, Rand DM (2010) Effects of selection and mutation on mitochondrial variation and inferences of historical population expansion in a Caribbean reef fish. *Molecular Phylogenetics and Evolution*, **57**, 821–828.

Hare MP, Nunney L, Schwartz MK *et al.* (2011) Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology*, **25**, 438–449.

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Herman JS, Searle JB (2011) Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **278**, 3601–3607.

Hey J (2010a) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Molecular Biology and Evolution*, **27**, 921–933.

Hey J (2010b) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Ho SYW, Lanfear R (2010) *Mito-communications*. *Mitochondrial DNA*, **21**, 33–35.

Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics*, **22**, 79–83.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources*, **11**, 423–434.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. *Systematic Biology*, **56**, 515–522.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE*, **3**, e1615.

Ho SYW, Lanfear R, Bromham L *et al.* (2011) Time-dependent rates of molecular evolution. *Molecular Ecology*, **20**, 3087–3101.

Hoban S, Bertorelle B, Gaggiotti OE (2012) Computer simulations: tools for population and evolutionary genetics. *Nature Reviews Genetics*, **13**, 110–122.

Hoelzel AR (1993) Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. *Current Opinion in Genetics and Development*, **3**, 891–895.

Hoelzel AR, Halley J, O'Brien SJ *et al.* (1993) Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *Heredity*, **84**, 443–449.

Hoffman JI, Grant SM, Forcada J, Phillips CD (2011) Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology*, **20**, 3989–4008.

Hope AG, Waltari E, Fedorov VB, Goropashnaya AV, Talbot SL, Cook JA (2011) Persistence and diversification of the Holarctic shrew, *Sorex tundrensis* (Family Soricidae), in response to climate change. *Molecular Ecology*, **20**, 4346–4370.

Horai S, Satta Y, Hayasaka K *et al.* (1992) Mans place in hominoidea revealed by mitochondrial DNA genealogy. *Journal of Molecular Evolution*, **35**, 32–43.

Howell N, Oostra RJ, Bolhuis PA *et al.* (2003a) Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. *American Journal of Human Genetics*, **72**, 1460–1469.

Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003b) The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. *American Journal of Human Genetics*, **72**, 659–670.

Howell N, Elson JL, Turnbull DM, Herrnstadt C (2004) African haplogroup L mtDNA sequences show violations of clock-like evolution. *Molecular Biology and Evolution*, **21**, 1843–1854.

Howell N, Kubacka I, Keers SM, Turnbull DM, Chinnery PF (2005) Co-segregation and heteroplasmy of two coding-region mtDNA mutations within a matrilineal pedigree. *Human Genetics*, **116**, 28–32.

Howell N, Elson JL, Howell C, Turnbull DM (2007) Relative rates of evolution in the coding and control regions of African mtDNAs. *Molecular Biology and Evolution*, **24**, 2213–2221.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxfords Surveys in Evolutionary Biology* (eds Futuyma DJ, Antonovics J), pp. 1–44. Oxford University Press, Oxford, UK.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature*, **425**, 175–178.

Irwin JA, Saunier JL, Niederstatter H *et al.* (2009) Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *Journal of Molecular Evolution*, **68**, 516–527.

Jackson JBC (2001) What was natural in the coastal oceans? *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5411–5418.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR (2009) Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder mysticeti). *Molecular Biology and Evolution*, **26**, 2427–2440.

Jezkova T, Olah-Hemmings V, Riddle BR (2011) Niche shifting in response to warming climate after the last glacial maximum: inference from genetic data and niche assessments in the chisel-toothed kangaroo rat (*Dipodomys microps*). *Global Change Biology*, **17**, 3486–3502.

Kemppainen P, Lindskog T, Butlin R, Johannesson K (2011) Intron sequences of arginine kinase in an intertidal snail suggest an ecotype-specific selective sweep and a gene duplication. *Heredity*, **106**, 808–816.

Kimmel M, Chakraborty R, Stivers DN, Deka R (1996) Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics*, **143**, 549–555.

Kimura M (1969) Number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Klutsch CFC, Seppala EH, Uhlen M, Lohi H, Savolainen P (2011) Segregation of point mutation heteroplasmy in the control region of dog mtDNA studied systematically in deep generation pedigrees. *International Journal of Legal Medicine*, **125**, 527–535.

Kokko H, Helle E, Lindström J, Ranta E, Sipilä T, Courchamp F (1999) Backcasting population sizes of ringed and grey seals in the Baltic and Lake Saiman during the 20th century. *Annales zoologici Fennici*, **36**, 65–73.

Koschinski S (2002) Current knowledge on harbour porpoises (*Phocoena phocoena*) in the Baltic Sea. *Ophelia*, **55**, 167–197.

Kuhner MK (2009) Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution*, **24**, 86–93.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kuhner MK, Yamato J, Beerli P *et al.* (2004) LAMARC. University of Washington, Available from http://evolution.gs.washington.edu/lamarc.html.

Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 803–808.

Lebreton JD, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals—a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

Leonard JA, Vila C, Wayne RK (2005) Legacy lost: genetic variability and population size of extirpated US grey wolves (*Canis lupus*). *Molecular Ecology*, **14**, 9–17.

Li CH, Riethoven JJM, Ma LB (2010) Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evolutionary Biology*, **10**, 90.

Liu JX, Tatarenkov A, Beacham TD , Gorbachev V, Wildes S, Avise JC (2011) Effects of Pleistocene climatic fluctuations on the phylogeographic and demographic histories of Pacific herring (*Clupea pallasii*). *Molecular Ecology*, **20**, 3879–3893.

Loogvali EL, Kivisild T, Margus T, Villems R (2009) Explaining the imperfection of the molecular clock of hominid mitochondria. *PLoS ONE*, **4**, e8260.

Lopes JS, Boessenkool S (2010) The use of approximate Bayesian computation in conservation genetics and its application in a case study on yellow-eyed penguins. *Conservation Genetics*, **11**, 421–433.

Lorenzen ED, Nogues-Bravo D, Orlando L *et al.* (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479**, 359–364.

Lotze HK, Worm B (2009) Historical baselines for large marine animals. *Trends in Ecology & Evolution*, **24**, 254–262.

Lotze HK, Coll M, Magera AM, Ward-Paige C, Airoldi L (2011) Recovery of marine animal populations and ecosystems. *Trends in Ecology & Evolution*, **26**, 595–605.

Lukacs PM, Burnham KP (2005) Review of capture-recapture methods applicable to noninvasive genetic sampling. *Molecular Ecology*, **14**, 3909–3919.

Lund MB, Olsen EM, Espeland SH, Stenseth NC (2011) Life history of fjord cod from the Skagerrak in the mid-2000s compared to 1905. *Marine Ecology Progress Series*, **424**, 169–174.

McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **360**, 1387–1393.

Meredith RW, Janecka JE, Gatesy J *et al.* (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science*, **334**, 521–524.

Millar CD, Dodd A, Anderson J *et al.* (2008) Mutation and evolutionary rates in adelie penguins from the Antarctic. *PLoS Genetics*, **4**, e1000209.

Moran PAP (1975) Wandering distribution and the electrophoretic profile. *Theoretical Population Biology*, **8**, 318–330.

Morgan K, O'Loughlin SM, Chen B *et al.* (2011) Comparative phylogeography reveals a shared impact of pleistocene envi-

ronmental change in shaping genetic diversity within nine Anopheles mosquito species across the Indo-Burma biodiversity hotspot. *Molecular Ecology*, **20**, 4533–4549.

Moritz C (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, **3**, 401–411.

Mourier T, Ho SYW, Gilbert MTP, Willerslev E, Orlando L (2012) Statistical guidelines for detecting past population shifts using ancient DNA. *Molecular Biology and Evolution*, **29**, 2241–2251.

Myers RA, Worm B (2005) Extinction, survival or recovery of large predatory fishes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **360**, 13–20.

Myers RA, Baum JK, Shepherd TD, Powers SP, Peterson CH (2007) Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science*, **315**, 1846–1850.

Navascues M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology*, **18**, 4390–4397.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science of the United States of America*, **76**, 5269–5273.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145–163.

Nicol S, Bowie A, Jarman S *et al.* (2010) Southern Ocean iron fertilization by baleen whales and Antarctic krill. *Fish and Fisheries*, **11**, 203–209.

Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*, **154**, 931–942.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

O'Connell MT, Shepherd TD, O'Connell AMU, Myers RA (2007) Long-term declines in two apex predators, bull sharks (*Carcharhinus leucas*) and alligator gar (*Atractosteus spatula*), in lake pontchartrain, an oligohaline estuary in southeastern Louisiana. *Estuaries and Coasts*, **30**, 567–574.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in canadian polar bears. *Molecular Ecology*, **4**, 347–354.

Palsbøll PJ, Clapham PJ, Mattila DK *et al.* (1995) Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Marine Ecology Progress Series*, **116**, 1–10.

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution*, **11**, 426–435.

Park L (2011) Effective population size of current human population. *Genetical Research*, **93**, 105–114.

Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, **21**, 3686–3687.

Pershing AJ, Christensen LB, Record NR, Sherwood GD, Stetson PB (2010) The impact of whaling on the ocean carbon cycle: why bigger was better. *PLoS ONE*, **5**, e12444.

Phillips CD, Trujillo RG, Gelatt TS *et al.* (2009) Assessing substitution patterns, rates and homoplasy at HVRI of Steller sea lions, *Eumetopias jubatus*. *Molecular Ecology*, **18**, 3379–3393.

Pinnegar JK, Engelhard GH (2008) The 'shifting baseline' phenomenon: a global perspective. *Reviews in Fish Biology and Fisheries*, **18**, 1–16.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

Polanski A, Kimmel M (2003) New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics*, **165**, 427–436.

Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, **181**, 711–719.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**, 291–300.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Punt AE, Friday N, Smith TD (2006) Reconciling data on the trends and abundance of North Atlantic humpback whales within a population modeling framework. *Journal of Cetacean Research and Management*, **8**, 145–160.

Rasmussen M, Guo XS, Wang Y *et al.* (2011) An aboriginal Australian genome reveals separate human dispersals into Asia. *Science*, **333**, 94–98.

Reeves RR, Mitchell E (1988) Current status of the gray whale, *Eschrichtius robustus*. *Canadian Field Naturalist*, **102**, 369–390.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rogers AR, Jorde LB (1995) Genetic evidence on modern human origins. *Human Biology*, **67**, 1–36.

Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science*, **301**, 508–510.

Ruegg KC, Anderson EC, Baker CS, Vant M, Jackson JA, Palumbi SR (2010) Are Antarctic minke whales unusually abundant because of 20th century whaling? *Molecular Ecology*, **19**, 281–291.

Santos C, Montiel R, Sierra B *et al.* (2005) Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Molecular Biology and Evolution*, **22**, 1490–1505.

Schwarz CJ, Seber GAF (1999) Estimating animal abundance: review III. *Statistical Science*, **14**, 427–456.

Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.

Shields GF, Wilson AC (1987) Calibration of mitochondrial DNA evolution in geese. *Journal of Molecular Evolution*, **24**, 212–217.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67–73.

Smith TD, Reeves RR (2003) Estimating American 19th Century catches of humpback whales in the West Indies and Cape Verde Islands. *Caribbean Journal of Science*, **39**, 286–297.

Soares P, Ermini L, Thomson N *et al.* (2009) Correcting for purifying selection: an improved human mitochondrial

molecular clock. *American Journal of Human Genetics*, **84**, 740–759.

Sremba AL, Hancock-Hanser B, Branch TA, LeDuc RL, Baker CS (2012) Circumpolar diversity and geographic differentiation of mtDNA in the critically endangered Antarctic blue whale (*Balaenoptera musculus intermedia*). *PLoS ONE*, **7**, e32579.

Subramanian S, Denver DR, Millar CD *et al.* (2009) High mitogenomic evolutionary rates and time dependency. *Trends in Genetics*, **25**, 482–486.

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tamura K, Kumar S (2002) Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Molecular Biology and Evolution*, **19**, 1727–1736.

Thomas L, Buckland ST, Rexstad EA *et al.* (2010) Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, **47**, 5–14.

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, **51**, 689–702.

Torroni A, Rengo C, Guida V *et al.* (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *American Journal of Human Genetics*, **69**, 1348–1356.

Vitalis R, Couvet D (2001) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics*, **157**, 911–925.

Waples RS (2002) Definition and estimation of effective population size in the conservation of endangered species. In: *Population Viability Analysis* (eds Beissinger SR, McCullough DR), pp. 147–168. The University of Chicago Press, Chicago, Illinois.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753–756.

Waples RS, England PR (2011) Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics*, **189**, 633–644.

Waples RS, 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. *Molecular Ecology*, **15**, 1419–1439.

Waples RS, Do C, Chopelet J (2011) Calculating N(e) and N(e)/N in age-structured populations: a hybrid Felsenstein-Hill approach. *Ecology*, **92**, 1513–1522.

Watterson GA (1984) Lines of descent and the coalescent. *Theoretical Population Biology*, **26**, 77–92.

Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, **182**, 1207–1218.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.

White GC, Burnham KP (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46** (Suppl), 120–138.

Willerslev E, Hansen AJ, Christensen B, Steffensen JP, Arctander P (1999) Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 8017–8021.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society Series A-Statistics in Society*, **166**, 155–188.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Wright S (1969) *The Theory of Gene Frequencies*. The University of Chicago Press, Chicago, Illinois.

Wu CH, Drummond AJ (2011) Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. *Genetics*, **188**, 151–164.

Yang ZH, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, **23**, 212–226.

The authors share an interest in the application of population genetic data and inference methods to conservation in addition to good food, wine and the Cal Bears.
