

INVITED REVIEW

Common misconceptions in molecular ecology: echoes of the modern synthesis

STEPHEN A. KARL,* R. J. TOONEN,* W. S. GRANT† and B. W. BOWEN*

**Hawai'i Institute of Marine Biology, University of Hawai'i, Mānoa, P.O. Box 1346, Kāne'ohe, HI 96744, USA, †Department of Biological Sciences, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508, USA*

Abstract

The field of molecular ecology has burgeoned into a large discipline spurred on by technical innovations that facilitate the rapid acquisition of large amounts of genotypic data, by the continuing development of theory to interpret results, and by the availability of computer programs to analyse data sets. As the discipline grows, however, misconceptions have become enshrined in the literature and are perpetuated by routine citations to other articles in molecular ecology. These misconceptions hamper a better understanding of the processes that influence genetic variation in natural populations and sometimes lead to erroneous conclusions. Here, we consider eight misconceptions commonly appearing in the literature: (i) some molecular markers are inherently better than other markers; (ii) mtDNA produces higher F_{ST} values than nDNA; (iii) estimated population coalescences are real; (iv) more data are always better; (v) one needs to do a Bayesian analysis; (vi) selective sweeps influence mtDNA data; (vii) equilibrium conditions are critical for estimating population parameters; and (viii) having better technology makes us smarter than our predecessors. This is clearly not an exhaustive list and many others can be added. It is, however, sufficient to illustrate why **we all need to be more critical of our own understanding of molecular ecology and to be suspicious of self-evident truths.**

Keywords: data interpretation, manuscript review, publishing research results, trends in molecular ecology

Received 16 December 2011; revision received 21 February 2012; accepted 7 March 2012

Introduction

In 1943 Julian Huxley published his seminal work 'Evolution: the modern synthesis' (Huxley 1943). Although some reviews were critical of certain aspects of the content and presentation, most were glowing (Hubbs 1943; Kimball 1943; Schmidt 1943). Huxley undertook this synthesis of the burgeoning field of evolution because isolation, miscommunication and misunderstanding were rampant in the sub-fields of biology that contributed most to evolutionary thought. He had hoped to explain how the contributions of theoretical population genetics, laboratory experiments and field research had resulted in a significant understanding of how evolution works. He also made a considerable effort to dispel

many commonly held misconceptions about evolution. In his review, Carl Hubbs (Hubbs 1943) felt compelled to point out that 'All biologists will profit by reading the book, and many professional workers sorely need to learn the lessons which it presents so clearly and penetratingly'. The primary factor underlying these misconceptions of evolution was that, although many sub-disciplines of biology were informing evolutionary thinking, many researchers within those sub-areas were not trained in evolutionary biology. They were incompletely aware of many of the mechanisms and processes of evolutionary biology. As such, many unfounded or poorly conceived and unsupported ideas about what is and is not important in evolutionary biology were being perpetuated.

The field of molecular ecology has reached a stage that might seem familiar to Huxley. We often encounter assertions in research articles, seminar presentations,

Correspondence: Stephen A. Karl, Fax: 808-236-7443; E-mail: skarl@hawaii.edu

reviews and comments from editors that seem reasonable on the surface, but prove to be either poorly supported or are misunderstandings of population genetic theory. These misconceptions arise from a complex mix of factors. Primary among them is inadequate training in population genetic and evolutionary theory. This is especially true for the many researchers from other fields that make contributions with little formal training in population genetics. Given the speed and relative ease with which molecular data can now be collected almost anyone can design, analyse and publish genetic data. The number of empirical studies in molecular ecology has exploded over the last few decades, since protein electrophoretic methods were first applied to population genetic studies in the late 1960s (e.g. Lewontin & Hubby 1966). The development of new technologies to detect genetic variation has allowed molecular ecologists to investigate problems that were intractable a few years ago. With the outsourcing of marker development, easy access to automated DNA sequencers, user-friendly software interfaces and ready access to large public databases, anyone with a computer can be a molecular ecologist, regardless of training. The situation is sometimes made worse by researchers, who after becoming familiar with a computer program, publish a few molecular ecological studies, become referees and begin to codify errant views in the discipline.

The field of molecular ecology encompasses numerous sub-disciplines, each with its own lineage of concepts. Misconceptions become enshrined in the literature when molecular ecologists fail to consider relevant concepts in other sub-disciplines. For example, in the sub-disciplines of phylogenetics, historical biogeography and phylogeography, molecular markers provide valuable insights into species' boundaries and the temporal framework of population divergence and dispersal. A goal of many of these studies is to understand the effects of past and present-day environmental variability on the genetic structures of populations, expressed by the dictum that 'earth and life evolve together' (Croizat 1964). While this premise was formulated to account for divergences between related taxa on different continents, it provides the motivation to search for causal relationships between paleoclimatic events (Lambeck *et al.* 2002; Jouzel *et al.* 2007) and genetic patterns within and among populations (e.g. Bermingham *et al.* 1997; Avise 2000). Misconceptions and errors can creep into molecular ecology studies, because of the failure to consider first-hand information in paleo-ecology and paleo-climatology.

Here, we identify eight common misconceptions that are frequently encountered in the broad field of molecular ecology. These misconceptions appear in print and are perpetuated because nonspecialists misapply

concepts in molecular ecology, especially population genetic theory. Indeed, a recent review of 137 mismatch analyses demonstrated that about half contained simple errors in calculating the age of a population expansion (Schenekar & Weiss 2011). Theoretical principles in the many sub-disciplines of molecular ecology are numerous and often complex, and it is easier to apply standard, widely used analyses than to dig into the original literature of related disciplines. We focus on common misconceptions that have repeatedly produced erroneous conclusions in the molecular ecology literature. The views presented in this review are incomplete, but hopefully will promote reflection and discussion.

Eight misconceptions

(i) *Some molecular markers are inherently better than others*

The field of molecular ecology is rife with simplistic statements that one class of marker is more sensitive to population structure than another class. This misconception is most sharply apparent with claims that mtDNA (or any haploid inherited organelle) will show population divergence first in recently divided populations due to higher levels of genetic drift, or that microsatellites will show divergence first due to high mutation rates and heterozygosities. Both can be true in individual circumstances, depending on a complex array of conditions that include genetic diversity, genetic effective population size (N_e ; i.e. the size of an idealized population that would experience the same amount of drift as the real population), mutation rate (μ) and migration characteristics, as well as sex-biased dispersal. No class of markers, however, is *a priori* more sensitive (i.e. is better able to detect population differentiation) under all conditions.

Under typical conditions of ongoing population divergence, mtDNA always has more power to detect population divergence than any single nuclear locus, but two or more polymorphic nuclear loci are expected to be more sensitive than mtDNA (Larsson *et al.* 2009). These findings are based on simulations in POWSIM, a software package that estimates the level of population divergence that can be detected with a given number of loci and sample size (Ryman & Palm 2006; Ryman *et al.* 2006). One important caveat is that diversities among markers in these simulations are held to be identical. A polymorphic mtDNA locus can have more power than a cluster of microsatellite loci depending on overall diversity in these markers, which will vary among species and evolutionary histories.

While it is clear that loci with low diversity have limited power to resolve differences, it is also true that

extremely high diversity can limit the power to detect population divergence. It is a mathematical certainty that high heterozygosity depresses F_{ST} values as demonstrated by Hedrick (1999). In addition, microsatellite loci can contain alleles that are identical in size (state) but not by descent (O'Reilly *et al.* 2004). The step-wise mutation model that predominates in microsatellite evolution produces a downward bias in estimates of population structure (by size homoplasy), relative to a marker evolving by the infinite allele model (Estoup *et al.* 2002). This effect will be most pronounced under scenarios of large population size ($N_e > 10^6$) and high mutation rate ($\mu > 10^{-3}$). The effect of high levels of allelic diversity on statistical power is not limited to microsatellites. For example, a survey of highly polymorphic mtDNA control region sequences in Pacific cod did not detect genetic partitions (Liu *et al.* 2010) that were apparent with less polymorphic mtDNA coding sequences (Canino *et al.* 2010).

Empirical data sets confirm that either mtDNA or microsatellites can detect population divergence not apparent in the other class of markers. Results for benthic (bottom dwelling) marine organisms are informative here because dispersal is accomplished almost exclusively through larvae, while juveniles and adults rarely move more than 1 km in a lifetime. Here, we can set aside concerns about sex-biased dispersal (and small population size in most cases), and ask how the inheritance of mtDNA and microsatellites shapes the magnitude of population divergence. **A review of the literature on reef fishes shows that, in some cases, mtDNA and not microsatellites will demonstrate more divergence and in other cases the opposite is true.** In an extreme example, a survey of microsatellite variation in the surgeonfish, *Zebrasoma flavescens*, detected seven populations and significant isolation by distance in the Hawaiian Archipelago ($F'_{SC} = 0.026$, $P < 0.001$), while the parallel mtDNA survey showed no significant differences ($\Phi_{SC} = 0.002$, $P = 0.38$; Eble *et al.* 2011). Clearly, both mtDNA and microsatellites can be more sensitive for detecting population divergence, and this is borne out in both theoretical (Larsson *et al.* 2009) and empirical studies (Eble *et al.* 2011).

It is now possible to interrogate tens of thousands of single nucleotide polymorphisms (SNPs) and to produce incredibly large data sets to search, for example, for genes under selection associated with adaptive traits (Hohenlohe *et al.* 2010). **While SNPs aptly facilitate genomic scans, they must be used cautiously to estimate gene flow, effective population size, genetic diversity and evolutionary mechanisms, because SNPs are often embedded in DNA segments with an unknown genetic background.** Methods that survey sequence variability, rather than single nucleotide positions, are

still recommended to answer many of the classical questions in population genetics that require estimates of genetic diversity, gene flow or historical and contemporary population sizes. Clearly it is not defensible to make blanket statements about the utility of one genetic marker over another (also see Schlötterer 2004 review). To evaluate the optimal markers for a particular study, much more than the mode of inheritance or mutability needs to be considered. Pertinent information will include locus diversity, available sample sizes, and the level of population divergence. Of course most of this information is only available once the laboratory aspect of the study has begun. However, the versatile molecular ecologist can adjust study design in response to these considerations. For example, a researcher who finds deep (or diagnostic) mtDNA divergences between populations might shift the nuclear DNA analysis from microsatellites to the less variable intron sequences, a more appropriate choice for molecular evolutionary separations.

(ii) mtDNA produces higher F_{ST} values than nDNA

The calculation of F_{ST} and its analogues (Φ_{ST} , F'_{ST} , G_{ST} , θ , R_{ST}) is surprisingly complex, and the appropriate choice of a F -statistic depends heavily on the level of genetic diversity (Waples & Gaggiotti 2006; Holsinger & Weir 2009; Bird *et al.* 2011). In particular, parametric F_{ST} has a downward bias in cases of high allelic diversity (typical of microsatellite loci). This can be corrected in a variety of ways (e.g. F'_{ST}) by calculating the upper limit for the F -statistics in each case, and scaling that range to fit the usual F -statistic range of 0.0–1.0 (Hedrick 1999; Meirmans & Hedrick 2011). Notably, Φ_{ST} , which takes sequence divergence into account, is usually larger than F_{ST} , except in special cases where deeply divergent lineages are distributed among populations, or where all haplotypes or alleles are equidistantly related (Bird *et al.* 2011).

During differentiation of two populations under ideal conditions (equal sex ratio, equal and low levels of migration, random mating within populations, no mutation and no selection), **simulations show that the ratio (R value) of mtDNA F_{ST} to nuclear F_{ST} ranges from $R = 1.0$ – 4.0** (Larsson *et al.* 2009). That means the F -statistics range from equality to four times higher in mtDNA. Examples of this range of R values are abundant in the literature (Table 1). During divergence between populations *without migration* both mtDNA and microsatellites theoretically start with $F_{ST} = 0.0$ at time 0, and both end with $F_{ST} = 1.0$ at equilibrium (typically after thousands of generations). It should be noted, however, that though the maximum F_{ST} is 1.0 at equilibrium, values at time 0 vary stochastically from

Table 1 Cases in which F -statistics for mtDNA are lower, equivalent, and higher than F -statistics for microsatellites (μ satDNA), ranked by R values (mtDNA F_{ST} /microsatellite F_{ST}). Note that R values far exceed the theoretical range of 1 to 4 in cases where sex-biased dispersal has been demonstrated. Some comparisons are made between regional groups (F_{CT}) rather than individual samples. The F_{ST} analogue is specified in each case. When comparing F -statistics, at least two biases are apparent: F_{ST} will usually be lower than Φ_{ST} for the same data set, and F_{ST} is biased downward relative to corrected F'_{ST} in data sets with high heterozygosity

Species	mtDNA	μ satDNA	R	References
Lower population structure in mtDNA relative to microsatellites*				
Smelt <i>Thaleichthys pacificus</i>	$F_{ST} = 0.023$	$F_{ST} = 0.045$	0.51	McLean & Taylor (2001)
Red grouse <i>Lagopus lagopus</i>	$\Phi_{ST} = 0.010$	$R_{ST} = 0.16$	0.63	Piertney <i>et al.</i> (2000)
Equivalent population structure in mtDNA and microsatellite loci				
Yellow Tang <i>Zebrasoma flavescens</i>	$\Phi_{CT} = 0.098$	$F'_{CT} = 0.116$	0.84	Eble <i>et al.</i> (2011)
Deepwater snapper <i>Pristipomoides filamentosus</i>	$\Phi_{ST} = 0.029$	$F'_{ST} = 0.029$	1.00	Gaither <i>et al.</i> (2011)
Caribou <i>Rangifer tarandus</i>	$F_{ST} = 0.128$	$F_{ST} = 0.127$	1.10	Cronin <i>et al.</i> (2005)
Higher population differentiation in mtDNA relative to microsatellite loci[†]				
Warbler <i>Dendroica caerulescens</i>	$F_{ST} = 0.019$	$F_{ST} = 0.011$	1.73	Davis <i>et al.</i> (2006)
Alligator snapping turtle <i>Macrochelys temminckii</i>	$\Phi_{ST} = 0.98$	$F'_{ST} = 0.43$	2.28	Roman <i>et al.</i> (1999), Echelle <i>et al.</i> (2010)
Sea otter <i>Enhydrus lutris</i>	$F_{ST} = 0.466$	$F_{ST} = 0.183$	2.55	Larson <i>et al.</i> (2002)
Lake whitefish <i>Coregonus clupeaformis</i>	$F_{ST} = 0.496$	$\theta = 0.161$	3.08	Lu <i>et al.</i> (2001)
Guanaco (llama) <i>Lama guanicoe</i>	$F_{ST} = 0.459$	$F_{ST} = 0.104$	4.41	Sarno <i>et al.</i> (2001)
Much higher population differentiation in mtDNA relative to microsatellite loci[‡]				
Humpback whale <i>Megaptera novaeangliae</i>	$\Phi_{ST} = 0.277$	$F_{ST} = 0.043$	6.44	Baker <i>et al.</i> (1998)
Hammerhead shark <i>Sphyrna lewini</i>	$\Phi_{ST} = 0.519$	$F_{ST} = 0.035$	14.80	Daly-Engel <i>et al.</i> (2012)
Sperm whale <i>Physeter macrocephalus</i>	$G_{ST} = 0.03$	$G_{ST} = 0.001$	30.00	Lyrholm <i>et al.</i> (1999)
Blacktip shark <i>Carcharhinus limbatus</i>	$\Phi_{ST} = 0.350$	$F_{ST} = 0.007$	50.00	Keeney <i>et al.</i> (2005)
Bechstein's bat <i>Myotis bechsteini</i>	$F_{ST} = 0.809$	$F_{ST} = 0.015$	53.90	Kerth <i>et al.</i> (2002)
Spectacled eider <i>Somateria fischeri</i>	$F_{CT} = 0.189$	$\theta = 0.001$	189.00	Scribner <i>et al.</i> (2001)
Loggerhead turtle <i>Caretta caretta</i>	$\Phi_{ST} = 0.42$	$F_{ST} = 0.002$	210.00	Bowen <i>et al.</i> (2005)

*Attributed to female-biased dispersal in the red grouse.

[†]Excluding cases of male-mediated dispersal.

[‡]Attributed to male-mediated dispersal.

0.0 due to sampling effects at the time of subpopulations division. At equilibrium, both markers (if adjusted for heterozygosity) yield equivalent F_{ST} values, and values during the intervening period will generally be higher for mtDNA, but the approach to equilibrium depends on the degree of population substructure, the local deme effective population size and migration rate between those demes (Whitlock & McCauley 1999). Simulations by Larsson *et al.* (2009) show that during

the march towards equilibrium, $R = 4.0$ initially, 1.6 in generation 200 and 1.0 in generation 1000.

As an illustration, the guanaco (wild llama) listed in Table 1 is an interesting case of a population on the island of Tierra del Fuego, isolated from mainland South America by a water barrier 8000 years ago (Sarno *et al.* 2001). This is a rare case of populations diverging in a known timeframe without migration, which would mean that the equilibrium value should

be $R = 1.0$. In contrast, the detected $R = 4.41$, indicate nonequilibrium conditions or other factors such as selection or strong drift influencing population divergence.

During population divergence with migration, simulations indicate that equilibrium values of F_{ST} for mtDNA are always higher than those for nuclear markers. Using a low but realistic migration rate of $m = 0.005$ (where m is the proportion of each population that receives migrants per generation), Larsson *et al.* (2009) calculate an equilibrium $F_{ST} = 0.66$ for mtDNA, and $F_{ST} = 0.33$ for nuclear loci. This yields $R = 2$; however, this ratio (and the disparity between F_{ST} values for the two classes of markers) rises towards $R = 4$ under scenarios of higher migration. The example here and the guanaco above underscore that straightforward theoretical expectations do not necessarily translate to the natural world, but do act as a touchstone for reasonable expectations and are guiding principles not binding regulations.

Sex-biased dispersal is an extreme form of divergence with migration, and this condition alters patterns of population subdivision and R ratios, as indicated by comparisons of uniparental and biparental markers (Karl *et al.* 1992; Bowen *et al.* 2005). Male dispersal predominates in many vertebrate groups, with higher divergence among populations recorded in mtDNA (Table 1). Female dispersal predominates in birds (Prugnolle & de Meeus 2002), and in at least one case yields higher F_{ST} in microsatellites than mtDNA ($R < 1$; Table 1). An interesting case of female-biased dispersal is recorded for the primate *Homo sapiens*, in which autosomal chromosomes, mtDNA and Y chromosomes yield estimates of genetic variance between continents of 8.8%, 12.5% and 52.7%, respectively (Seielstad *et al.* 1998). In the anadromous fish *Thaleichthys pacificus* from the northeast Pacific, the microsatellite value is $F_{ST} = 0.045$, while the corresponding mtDNA value is $F_{ST} = 0.023$ ($R = 0.51$ in Table 1; McLean & Taylor 2001). Clearly, F_{ST} values from either mtDNA or microsatellites can be higher, depending on a complex set of conditions. The haploid inheritance of mtDNA (and other organelles) confers higher F_{ST} values under most conditions, but both theoretical and empirical studies show that this is not invariably true.

(iii) Estimated population coalescences are real

MtDNA genealogies are commonly used to infer historical demographies with coalescence theory (Kingman 1982), implemented in sequence mismatch analysis (Rogers & Harpending 1992) and Bayesian skyline plots (BSP; Drummond & Rambaut 2007), among other methods (Hey & Nielsen 2004). These methods produce

estimates of compound parameters that include effective population size and mutation rate. Estimates of mutation rate are needed to extract the population variables and to date population events. However, several sources of error, including sample size and estimates of mutation rate, can seriously compromise the accuracies of coalescence-based analyses to infer population histories.

To illustrate some of these errors, we use coalescence simulations of nonrecombining DNA sequences under a population history of recent population growth that is typical for marine species (Box 1). These simulations show variability in the gene genealogies within a population and times to most recent common ancestor (TMRCA) for two sample sizes (Figs 1a and 2a). TMRCA among replicate genealogies varied by a factor of two, and shapes of the genealogies varied considerably among replicates, even for the same sample size. In practice, the distributions of mutations along branches can then be used to reconstruct a genealogy (Figs 1b and 2b). In addition to coalescent variability, an observed DNA gene genealogy reflects only one realization of many possible mutation histories. In our simulations, mutation trees largely captured deep partitions in the coalescent trees, but did not always resolve relationships in the upper (younger) part of the trees. The variability among realized DNA trees can also be seen in

Box 1. Coalescence modelling

Coalescence simulations of DNA genealogies are made in two steps (Hudson 1990). First, a coalescence tree depicting the genealogical relationships among individuals in a sample is created by moving backward in time. At each generation, the model assigns a common ancestor to two individuals or groups based on effective population size. Since coalescences between lineages occur more rapidly in small populations, genealogies in small populations are shallower than in large populations. Coalescences between lineages continue each generation until the most recent common ancestor (MRCA) is reached at the base of the genealogy.

In the second step, mutations are placed on the genealogy in the forward direction beginning with the MRCA. The amount of detail in the genealogy captured by mutation depends on the mutation rate. A small mutation rate may show deep partitions in the tree, but may fail to show recent population events. A large mutation rate may resolve the upper branches and twigs in the tree, but not the deep history of the population.

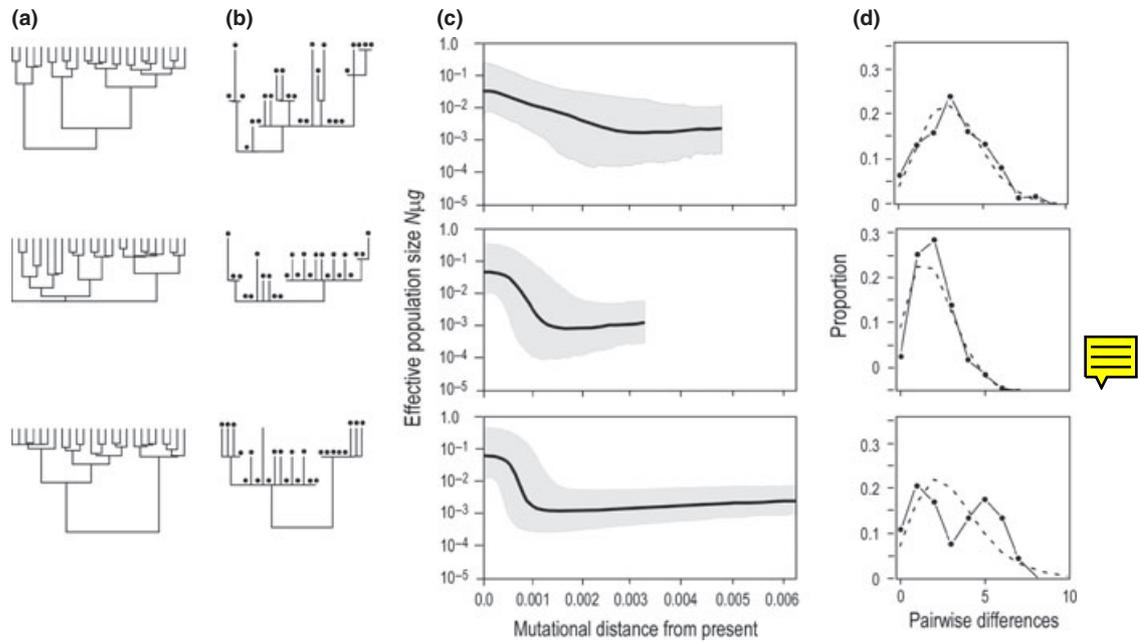


Fig. 1 Coalescence genealogies (a), mutation trees (b), Bayesian skyline plots (c) and mismatch distributions (d) for three coalescence simulation with sample size $n = 25$ drawn from a population that experienced a ‘knife-edge’ growth in size $N_e = 1\ 000$ to $1\ 000\ 000$ at 250 generations in the past (See supplemental information for details of simulations).

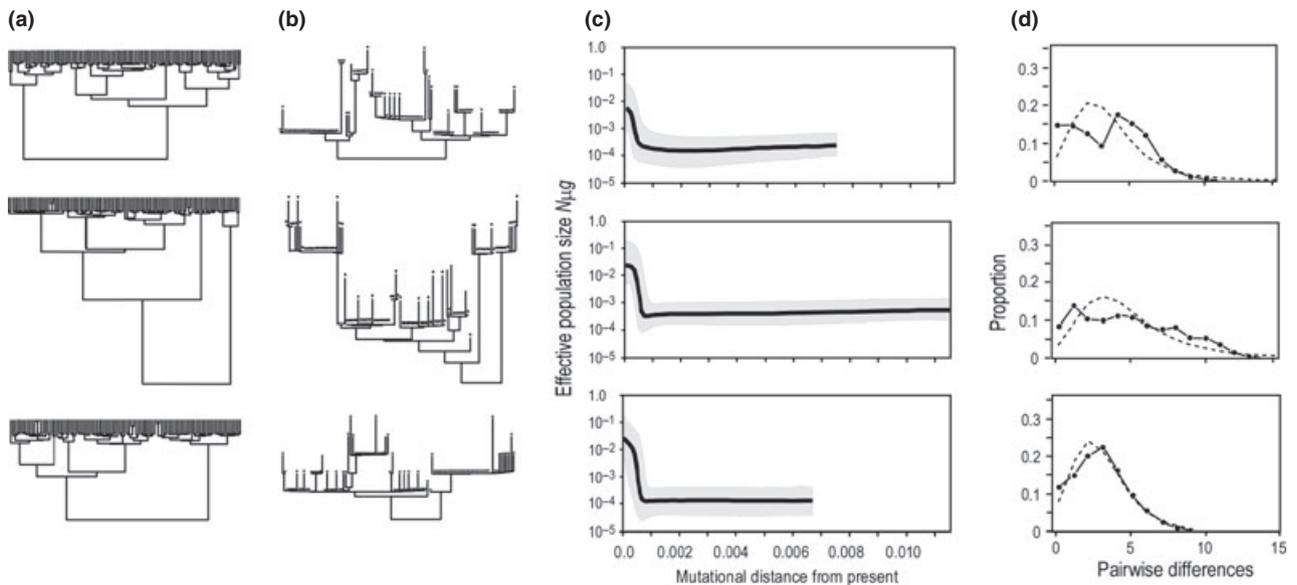


Fig. 2 Coalescence trees (a), one realization of a mutation tree (b), Bayesian skyline plots (c) and (d) observed (closed circles) and expected (expanding population) mismatch distributions for three coalescence simulations with sample size $n = 100$. Demographic model and explanation of figures as in Fig. 1.

the contrasting shapes of Bayesian skyline plots (BSPs; Figs 1c and 2c) and mismatch distributions (Figs 1d and 2d). Remarkably, these results were generated with the same demographic and mutation models.

These simulations show how coalescent and mutational randomness conspire to produce a variety of

mtDNA genealogies for the same population history (Rosenberg & Nordborg 2002). However, molecular ecologists do not always appreciate that a single molecular genealogy perhaps produced by months of field and laboratory work, represents only one of an infinite number of possible coalescent and mutational realiza-

tions. In the hands of most molecular ecologists, data sets producing contrasting BSPs and mismatch distributions generally prompt different interpretations. For example, small differences in shapes of BSPs were used to argue alternative hypothesis of population colonization and expansion (e.g. peopling of the Americas: Kitchen *et al.* 2008; Fagundes *et al.* 2008). When samples are difficult to collect or to sequence, we often attempt to maximize our efforts by resorting to batteries of statistical tests. The pitfall of this approach, however, is the temptation to over-interpret results.

Another source of error is inaccurate estimates of mutation rate (μ) to calibrate a molecular clock. In marine studies, the closure of the Panama Seaway in the late Pliocene (Marko 2002; Coates *et al.* 2005) and the opening of Bering Strait in the early Pliocene (Verhoeven *et al.* 2011) are commonly used to calibrate μ . When an internal calibration is unavailable, researchers use a proxy calibration based on other taxa, or a 'universal' molecular clock rate (e.g. Bowen & Grant 1997). These phylogenetically derived mutations rates, however, appear to overestimate the ages of phylogeographical events inscribed in genetic data, sometimes by an order of magnitude (Ho *et al.* 2005, 2008; Crandall *et al.* 2012). As a result, BSPs and mismatch analyses in many studies appear to indicate population expansions during glacial maxima (Canino *et al.* 2010; Liu *et al.* 2010, 2011; Stamatis *et al.* 2004; Strasser & Barber 2009; Pérez-Losada *et al.* 2007; Marko & Moran 2009; Carr & Marsall 2008; Hoarau *et al.* 2007; among many others). These scenarios are unlikely, because marine populations contract and expand in response to decadal environmental shifts (Perry *et al.* 2005) and larger environmental disturbances are expected to have correspondingly larger effects on population abundances and distributions.

One possible explanation for inaccurate molecular clocks is that mutation rates may be 'time dependent' (Ho *et al.* 2005). Calibrations based on recent divergences between taxa show much larger mutation rates than calibrations based on ancient phylogenetic divergences for birds (Ho *et al.* 2005), primates (Ho *et al.* 2005, but see Emerson 2007) and marine invertebrates (Crandall *et al.* 2012). The apparent elevation in mutation rate in recently diverged populations may be due to several factors, without having to invoke changes in the instantaneous rate of mutation. One source of error stems from the failure to account for polymorphisms in an ancestral population before it split into isolated populations destined to become new species (Hickerson *et al.* 2003; Charlesworth 2010). This effect is magnified in large populations, such as those in many marine species, and with the use of recent separation times to calibrate the molecular clock. Back-

ground selection on slightly deleterious alleles (Ho & Larson 2006; but see Peterson & Masel 2009) and balancing selection (Charlesworth 2010) may also contribute to apparent elevated mutation rates in recent divergences.

In many cases, the incorrect dating of phylogeographic events may be an artefact of a particular analytical method (e.g. mismatch analysis or BSPs) that does not distinguish between different histories of gene lineages in a sequence data set. For example, mtDNA data sets often consist of shallow, star-shaped lineages connected by deeper separations. When the star-shaped lineages are examined individually, the use of 'standard' phylogenetically derived estimates of mutation rate yields reasonable temporal estimates of recent population events (e.g. Saillard *et al.* 2000). Appropriate 'apparent' mutation rates for some methods of analysis can be estimated empirically with the analytical method itself. For example, Crandall *et al.* (2012) used BSPs to estimate population expansion dates in three marine species inhabiting the Sunda Shelf by reasoning that an expansion could only have occurred after the last glacial maximum (LGM), when rising sea levels submerged the shelf. Alternatively, Grant & Cheng (in press) simulated mtDNA sequences under a demographic model constructed from Pleistocene temperatures (Jouzel *et al.* 2007) to date the expansion of red king crab populations in the North Pacific (Fig. 3).

In addition to providing an empirical mutation rate, our simulations demonstrate several features of coalescence analysis that can lead to erroneous inferences (Fig. 4). First, a putative stable population history preceding a recent population expansion (as reported in many cases) may be an artefact of coalescence analysis. Second, only the most recent episode of rapid population growth can be detected, even if the populations experienced several periods of growth and decline. Population declines during the LGM may not be severe enough to lower genetic diversities, but are sufficient to erase information about previous population swings. This loss of information results in a flat population curve that is often erroneously interpreted as population stability over much of the Pleistocene. Third, a spike in population size is associated with warming after the last glacial maximum 18 000–20 000 years ago. However, the use of the wrong mutation rate (Ho *et al.* 2011) or inattention to ancestral polymorphisms (Hickerson *et al.* 2003) can place this almost universal signal of population growth in a previous interglacial period or even at a glacial maximum. Molecular ecologists often test phylogeographic models with standard computer programs and with standard estimates of mutation rate without appreciating the pitfalls of coalescence-based analyses. Though coalescence-based

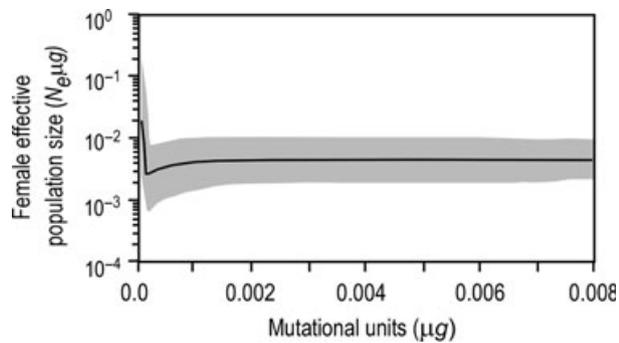


Fig. 3 Bayesian skyline plots (BSPs) based on mitochondrial cytochrome oxidase I sequences (bp = 665) in red king crabs ($n = 551$) in the central North Pacific and Bering Sea. Historical apparent effective population size (thick line) is bracketed by the 95% highest probability density (grey). The BSP was constructed with BEAST 1.6 (Drummond & Rambaut 2007) under the TrN (Tamura & Nei 1993) model of nucleotide substitution, ten piecewise linear intervals and a strict molecular clock. A MCMC run of 400 million steps yielded an effective sample sizes (ESS) of at least 200.

analyses are valuable and informative, their estimation and interpretation need to be very carefully considered.

(iv) *More data are always better*

Molecular ecologists live in exciting times. Not only has the availability of molecular tools considerably increased in number and ease of use, but analytical approaches have kept pace. With such sequencing methods as Roche 454 pyro- and Illumina sequencing and Bayesian algorithms to analyse data, many questions can be addressed that were previously impossible or were possible only with model organisms. For example, Hohenlohe *et al.* (2010) used 45 000 SNPs in 20 threespine sticklebacks from each of five locations (two oceanic

and three freshwater forms) and found that several loci were likely under selection and responsible for phenotypic differences among groups. Other researchers used entire mitochondrial genomes (~16 700 bp) to address evolutionary questions such as the origins of freshwater fishes (Nakatani *et al.* 2011). Neither of these studies could have been conducted 15 years ago. While researchers now have the ability to collect and analyse large parts of the genome quickly, are these large amounts of data helping to answer classic questions? The answer is surprisingly complex.

To determine how much data to collect, one must consider how much data are needed to produce robust conclusions. Will large amounts of data resolve questions that were not answered with smaller data sets because of weak signal or too little power? In the case of sticklebacks, only a large amount of data could support the conclusions of the study. Here, the question was which genes are likely responsible for the evolution of body forms in sticklebacks. A large data set of 45 000 SNPs greatly enhanced the chances that some of these markers would be linked to regions in the genome responsible for phenotypic differences. Though the conclusions are tentative, they provide a strong foundation for unravelling the genetic basis of adaptive mechanisms.

The study of the systematics of flightless (ratite) birds provides a contrasting example. Traditionally, both morphological and molecular studies indicated a monophyletic ratite grouping, including Cassowary, Emu, Kiwi, Ostrich, Rheas and Moa, but excluded the flighted sister taxon, the Tinamous (Prager *et al.* 1976; Sibley & Ahlquist 1990). Two studies using complete or near complete sequences of the mtDNA genome supported this model (Cooper *et al.* 2001; Haddrath & Baker 2001). Two studies of at least 19 nuclear DNA sequences from the ratites and Tinamous indicated that Tinamous clus-

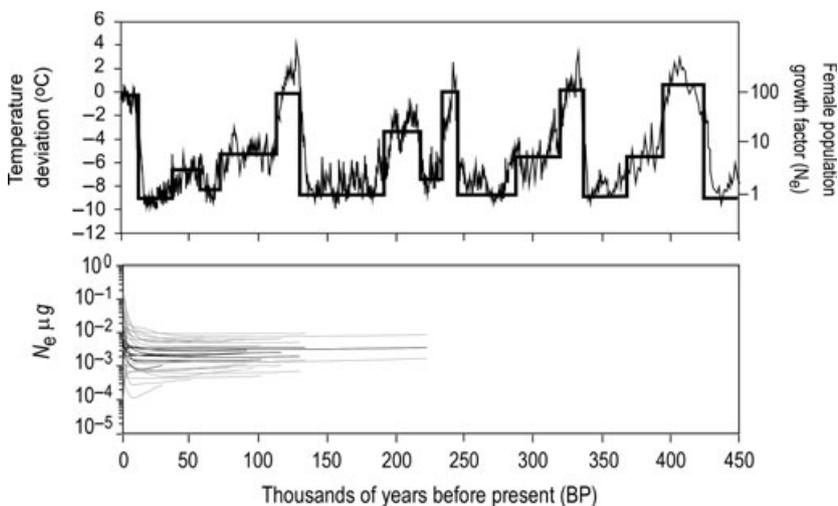


Fig. 4 Ten replicate simulations (bold lines) of historical demography in red king crab to illustrate the extent that coalescence analysis of mtDNA sequences captures population size histories over the last several ice-age cycles. Grey lines enclose 95% highest probability densities around estimates of historical demography.

tered within the ratite group and was a sister taxon to a Cassowary-Emu-Kiwi lineage (Hackett *et al.* 2008; Harshman *et al.* 2008) implying that ratites are paraphyletic. Phillips *et al.* (2010) undertook a second whole mtDNA study to resolve this problem and the new results supported the ratite paraphyly found with nDNA.

Did more data result in different conclusions? In some ways they may have, but in other ways probably not. In the nuclear studies, the systematic relationships among the taxa were estimated from multiple, unlinked loci. Basing phylogenetic relationships on multiple markers is generally a more robust approach, because it dilutes the vagaries of single-marker evolution (Felsenstein 2006). For the nDNA analysis, more loci added useful information. As a nonrecombining genome, however, the entire mtDNA molecule can be considered a single-locus and the mtDNA tree may not reflect a species-level phylogeny (Avise 1994). Though Phillips *et al.* (2010) included more mtDNA data (i.e. two additional kiwi species), they also used the same sequences from the previous mtDNA studies (Cooper *et al.* 2001; Haddrath & Baker 2001). The new kiwi sequences clustered with the old kiwi sequences, so the new data, clearly, did not alter the conclusion. A major difference among the studies, however, was that Phillips *et al.* (2010) used different analytical approaches and a different DNA mutation model. An underlying difficulty is that these birds likely radiated rapidly in the ancient past, so the evolutionary signal of relationship in mtDNA at deeper nodes has largely been lost. Hence, an absolute resolution of this debate is unlikely with mtDNA. It is comforting, however, to know that with new analyses, mtDNA can be concordant with the results from nDNA. Overall, it is important to keep in mind that some evolutionary questions cannot be definitively answered with DNA data because the event took place too long ago, or because several lineages diverged over the same timeframe, or both. **The important considerations when robust conclusions are lacking are the sensitivity and power of the data. When reporting results where it is clear that the markers had little sensitivity (i.e. were not variable enough) and low power (e.g. few loci were used), it is appropriate to acknowledge that more data might change or refine the conclusions. If, however, all analyses and markers strongly indicate the same result, adding more data simply to reach some idealized number of loci or sequence length is unlikely to add further insight.**

(v) *One needs to do a Bayesian analysis*

Concomitant with the huge volume of data that can be generated in a relatively short period of time, analytical

approaches have dramatically increased in number and approach. We acknowledge that none of the authors has thorough training in mathematics or statistics and we certainly do not want to add more misconceptions to the literature. We can, however, relate some of the pitfalls to new and intellectually compelling analytical methods. One of the first computer programs to analyse population genetic data was BIOSYS-1 (Swofford & Selander 1981). It is a straightforward FORTRAN program that provides the basic analyses of genetic data [e.g. fit to Hardy–Weinberg expectations, similarity and distance measures, Wright's *F*-statistics (Wright 1943), etc.]. A citation report from The Web of Knowledge (<http://apps.webofknowledge.com>) shows a peak in citations in 1996 (180) with a gradual drop to 16 in 2011 (Fig. 5). A newer program, GENEPOP (Raymond & Rousset 1995), shows a similar pattern with a gradual rise and fall, peaking in 2009. There are two differences between the pattern of citation for GENEPOP and BIOSYS-1. Notably, BIOSYS received 180 citations at its peak and a total of 2 205 citations, whereas the peak GENEPOP citation number was 909 in 2009 and a total of 7 740 as of 31 December 2011 (Table 2). There are clearly many more publications dealing with population genetic data now than in the heyday of BIOSIS. It is also interesting to note that both BIOSYS and GENEPOP peaked in citations 14 years after they were introduced. Though citations for several other analysis programs have shown a decline in 2010 or 2011, it is still too early to tell whether these trends will continue. Logically, it seems reasonable that the trend seen for BIOSYS will be replayed as new techniques and approaches are developed. The point is that, there has always been some new, hot analytical method and it is this method that is generally believed to be the best

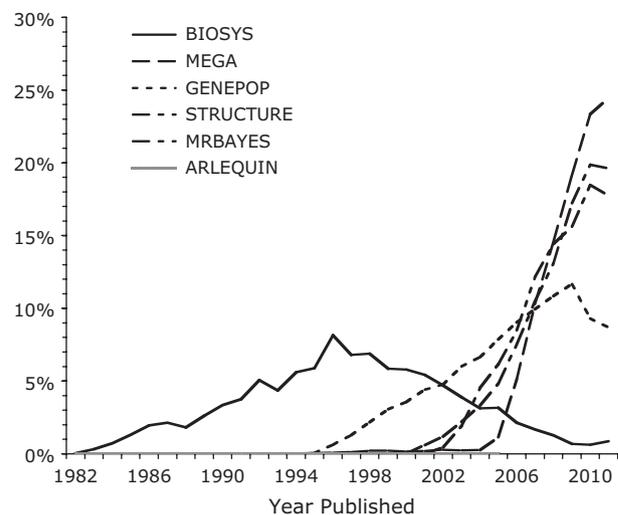


Fig. 5 Percent of total citations to date (31 December 2011) for a variety of population genetic analytical programs.

Table 2 Citation data for several commonly used genetic analyses programs. Data were obtained from the Web of Science searching for the publications associated with the programs and includes the year since published to 31 December 2011

Program	Year published*	Total no. of citations	Average no. of citations per year
BIOSYS	1981	2 205	73.50
MEGA†	1994	18 759	1 042.17
GENEPOP	1995	7 740	455.29
STRUCTURE	2000	5 104	425.33
MRBAYES‡	2001	14 836	1 348.73
ARLEQUIN	2005	4 189	698.17

*When there are multiple versions of a program, only the earliest data is given.

†There are five versions published in 1994, 2001, 2004, 2007 and 2011. Data include citation to all versions.

‡There are two versions published in 2001 and 2003. Data include citation to all versions.

approach. Of concern, however, is that reviewers and editors often criticize a manuscript because the authors did not use the latest approach regardless of the robustness of their conclusions. In addition, authors may want to use whatever the hottest program is regardless of their understanding of the mathematical approach and the appropriateness of the method.

An old, but revived analytical approach (Bayes 1763) has recently been applied to population genetic and phylogenetic analyses. These are Bayesian approaches that estimate the distribution of a parameter based on the collected data. One of the more widely used programs for estimating population subdivision is STRUCTURE (Pritchard *et al.* 2000) and for phylogeny reconstruction is MRBAYES3 (Ronquist & Huelsenbeck 2003). The literature citation patterns for these programs are similar to BIOSYS and GENEPOP (Fig. 5). As with GENEPOP, it is too soon to tell for how long these trends will continue. Definitely, these are useful and informative computer programs. The fundamental question here is, does a Bayesian approach provide more or deeper insight than other approaches?

One of the strengths of a Bayesian method is that, several types of data can be combined into a single analysis and multiple parameters can be estimated simultaneously. It is intellectually compelling to include as much information as is known when trying to reconstruct a complex event. Surprisingly, however, published genetic studies often use uninformative priors (e.g. uniform or flat) and include no other information beyond the genetic data. We are not suggesting that the use of uninformative priors is resulting in erroneous results, just that the true power of a Bayesian analysis lies in the ability to bring additional information to the

estimation. More importantly, however, eschewing informative priors causes a Bayesian analysis to converge on a likelihood analysis (Dale 1999). Notably, the criteria for priors are highly debated. For the most part, if there are sufficient data and the underlying signal is strong, Bayesian analyses are robust to the choice of priors (King *et al.* 2010). That is, if the analytical results are highly significant and the data uniformly indicate the same solution, then a Bayesian analysis with uninformative priors is likely to result in the correct solution. If, however, the data are few or not particularly informative, choosing an inappropriate flat prior can adversely affect the outcome (King *et al.* 2010) and may simply result in returning the prior value for the parameter being estimated. It is also probably true that if the data are sufficiently informative to remove the importance of the prior, a non Bayesian analysis is likely to produce the same result. Another limitation to Bayesian (as well as likelihood) approaches is that they can take a very long time to run, especially with large data sets. As such, they are rarely, rigorously tested using scenarios mirroring natural populations. Even more so, when they are tested (Faubet *et al.* 2007), there are many realistic conditions under which they perform poorly. Even more troubling is that incorrect answers can be associated with high confidences (i.e. posterior probability). When suggesting or evaluating a method of data analysis, it is important to assess how strong the result is and determine whether there is a benefit to a different approach. In many cases in molecular ecology, the information needed to choose appropriate priors for a Bayesian analysis is mostly lacking. Though Bayesian analyses are clearly powerful and can, at times, provide a solution where other approaches cannot, they are not always the best approach.

(vi) Selective sweeps influence mtDNA data

A selective sweep is the process by which a beneficial mutation increases in frequency relative to other alleles in the population and, all else being equal, ultimately becomes the only allele in the population (i.e. fixed). One outcome of a selective sweep is that loci linked to that mutant allele also increases in frequency, a process called genetic hitchhiking (Kaplan *et al.* 1989; Braverman *et al.* 1995). Hence, in the case of strong selection, the rapid fixation of a *de novo* beneficial mutation can eliminate genetic variation in a portion of the genome (Maynard Smith & Haigh 1974; Nielsen 2005). Alternatively, changing selection pressures can favour a previously neutral allele, which would also purge genetic variation from the population but not to the same extent as a *de novo* one with the same selection coefficient value.

Because the animal mitochondrial genome typically does not undergo recombination (Birky 2001), any single codon affected by selection will produce a hitchhiking effect for the entire molecule, in principle making mtDNA particularly sensitive to selective sweeps. Ballard & Whitlock (2004) reviewed the evidence for mechanisms of selective sweeps and a suite of studies documenting selective sweeps in animals. One particularly clear example is the spectacular impacts of *Wolbachia* (a maternally inherited α -proteobacteria that causes a variety of reproductive abnormalities in the host) that can result in only a single haplotype dominating an entire population (e.g. Turelli & Hoffmann 1991; Nurminsky *et al.* 1998). For example, in *Drosophila simulans*, *Wolbachia* infection induces cytoplasmic incompatibility such that an infected male, mating with a female that does not carry that same strain of *Wolbachia* or is uninfected, will produce a reduced number of offspring or be effectively sterile (Turelli & Hoffmann 1991). This is clearly strong selection pressure for the fixation of a single strain of *Wolbachia*. Due to the potential role of hitchhiking in shaping mtDNA diversity, selective sweeps are often invoked when addressing a surprising or counter-intuitive result. It is the argument most frequently used to downplay the value of single-locus mtDNA studies, but how often is it really happening?

In some cases, conflicting patterns inferred from nDNA and mtDNA are interpreted as evidence of a selective sweep (e.g. Houliston & Olson 2006; Linnen & Farrell 2007), whereas in others it is interpreted as evidence of introgression, some demographic historic impact, sex-biased dispersal (e.g. Fay & Wu 1999; Rokas *et al.* 2001; Bowen *et al.* 2005; Gompert *et al.* 2006) or some combination of these events (e.g. Rato *et al.* 2010). The majority of studies use statistical tests of linkage disequilibrium around the targets of selection to detect a selective sweep (Kim & Stephan 2002; Kim & Nielsen 2004; Nielsen 2005). Essentially, these tests examine whether a given haplotype is overrepresented in the population. Under neutral evolution, genetic diversity in a population is expected to be a function of the product of the genetically effective size (N_e) and the mutation rate (μ).

Even though selective sweeps are often invoked, the number of studies reporting empirical evidence for them is surprisingly small (reviewed by Ballard & Whitlock 2004; Dowling *et al.* 2008). Among the most commonly cited support for the wide-spread action of selective sweeps on mtDNA is the work of Bazin *et al.* (2006) which showed that mtDNA diversity does not follow intuitive predictions about population size in a survey of approximately 3 000 animals. Bazin *et al.* (2006) showed that nuclear but not mtDNA variability generally fit predictions of levels of genetic diversity

based on population sizes, which are expected to be larger for invertebrates than vertebrates, marine than terrestrial, and smaller than larger organisms. The poor fit of mtDNA diversity to neutral expectations based on population sizes was explained by frequent selective sweeps, and the authors conclude that '...recurrent adaptive evolution challeng[es] the neutral theory of molecular evolution and question[s] the relevance of mtDNA in biodiversity and conservation studies' (Bazin *et al.* 2006). In response, Mulligan *et al.* (2006) use the same methodology to show that the expected correlation between nuclear and mitochondrial DNA diversity and population size is robust in the well-studied eutherian (placental) mammals. Wares *et al.* (2006) further point out that the neutrality index (NI) developed by Rand & Kann (1996) and used by Bazin *et al.* (2006) is appropriate for only closely related taxa such as the eutherian mammals, and the test is biased to find selection between more distantly related organisms. Wares *et al.* (2006) finally point out that the comparative paucity of exhaustive invertebrate phylogenies forces more distant outgroup comparisons in the analysis of Bazin *et al.* (2006). The suite of responses to Bazin *et al.* (2006) argues that the observed pattern provides only indirect inference of selective sweeps in animal mitochondria. Likewise, in a survey of 162 well-studied fish species for which contemporary abundance can be accurately estimated, McCusker & Bentzen (2010) found a strong association between abundance and measures of genetic diversity for both mtDNA and microsatellites. They conclude that results 'generally conformed to neutral expectations' for these markers, and found no evidence of selective sweeps for either nuclear or mitochondrial markers.

If selective sweeps are a common and ubiquitous process then why is mtDNA variation roughly three-fold higher than nuclear variation in the Bazin *et al.* (2006) study? Clearly the subject of what processes drive variation in mtDNA among natural populations is complex and incompletely understood (reviewed by Ballard & Whitlock 2004; Dowling *et al.* 2008; see also Theisen *et al.* 2008). **Any simple generalization is indefensible with the data at hand; however, the abundance of mtDNA diversity in natural populations indicates that selective sweeps of the mitochondrial genome are rare.**

(vii) *Equilibrium conditions are critical for estimating population parameters*

Many of the analyses and theoretical principles in molecular ecology assume, explicitly or implicitly, that the population under consideration is in equilibrium for the four factors that change allele frequencies: mutation,

drift, migration and selection. Population size is not changing, so the rate of drift is the same as it was generations ago. Migration barriers between two subpopulations have not recently been removed or established, and the rate and direction of migration is not changing. One reason for the assumption of equilibrium is simple; genetic studies are mostly single slices in time, but draw conclusions about what happened in the past or will happen in the future. For example, a population experiencing a recent bottleneck is likely to retain most of the ancestral heterozygosity. Low-frequency alleles are lost in a bottleneck but they contribute little to the overall heterozygosity levels, and only extreme and sustained bottlenecks will result in extensive inbreeding (Nei *et al.* 1975). If we assess a population soon after a bottleneck, we would estimate a genetically effective population size much larger than it would be at equilibrium because the expected loss of heterozygosity due to inbreeding requires a sustained bottleneck. The unfortunate reality is that the evolutionary forces acting on populations are always changing, and it is likely that few natural populations are ever in complete equilibrium. Should we then not undertake analyses that assume equilibrium? **Though we urge caution, we think that avoiding analyses that assume equilibrium is an extreme view.**

Natural populations are distributed over geographic space with varying degrees of gene flow connecting subregions. Those subregions where gene flow is high are generally considered panmictic (i.e. a single population). Subregions connected by limited gene flow will, over evolutionary time, differentiate in allele frequencies (assuming no selection). There are several ways to estimate the magnitude of differentiation among subpopulations (e.g. F'_{ST} , G'_{ST} , etc.) and these can be very useful in describing the genetic architecture of a species. One important assumption in all of these parameters, however, is that the populations under consideration have reached genetic equilibrium. If natural populations are not in equilibrium, is it useful to try to estimate the magnitude of differentiation?

The answer to this question depends on how far out of equilibrium the population is and the effect of this deviation on population parameter estimation. Unfortunately, neither of these have easy answers. On the one hand, if populations are never in equilibrium due to physical and biological perturbations and deviation from equilibrium has a significant affect, then analyses assuming equilibrium should be avoided. No hard and fast rule is applicable, because some population variables (e.g. F'_{ST}) can return to equilibrium quickly after significant deviations (Crow & Aoki 1984; Birky *et al.* 1989; Whitlock & McCauley 1999), whereas others (mismatch distribution) may not, and the rate of approach

to equilibrium often depends on other parameters such as mutation rates and N_e . In contrast, if natural populations are never in equilibrium, the equilibrium value is of theoretical not empirical or practical concern. Presumably, we are estimating a parameter to gain insight into a real population. If the real population never attains a theoretical ideal, the measurement taken out of equilibrium is more reflective of the actual population. Attaining equilibrium can take 10 000s of generations (Birky *et al.* 1989), depending on rates of migration, N_e , mutation and drift. Even so, movement to equilibrium follows an asymptotic curve with the largest change in the first 100's of generations, followed by a long, gradual approach to true equilibrium (Wright 1965; Whitlock & McCauley 1999). Hence a population will reach a state close to equilibrium fairly quickly and retain this status for most of the march towards equilibrium (Slatkin 1993).

There may be some clues as to how close a population is to equilibrium. For example, the green crab, *Carcinus maenas*, is a highly successful aquatic invasive species having established populations in temperate regions of all continents during the last several centuries of ship traffic. Darling *et al.* (2008) use genetic analyses to reveal that the Atlantic US coastal population was introduced from Europe and subsequently spread to the west coast of North America. Samples from the east and west coast are genetically indistinguishable. Discarding the possibility that east and west coast populations represent a panmictic group, genetic data alone yields an incorrect conclusion about population structure, because the cessation of migration between the two groups is too recent, and the populations have not reached migration-drift equilibrium. Alternatively, many North American species were extirpated from their northern ranges during Pleistocene glaciations. For example, the chestnut-backed chickadee (*Poecile rufescens*) was likely limited to the southern part of its western North American range until the northward retreat of the Cordilleran glacier (~12 500 years ago). Results of genetic analysis (Burg *et al.* 2006) indicated population differentiation among many but not all of the sampled populations. Although the age of the northern recolonization is unknown, there likely have been 1000s of generations since that event. In this case, it is unlikely that nonequilibrium conditions are adversely affecting the results. The chestnut-backed chickadee may not be at migration-drift equilibrium, but it is likely closer to equilibrium than the green crab. **The point here is that when considering results assuming equilibrium, it is prudent to ponder two related questions: 1) are the results consistent among tests and with what else is known about the species under consideration; and 2) are the inferences from these analyses**



couched with proper caveats and alternative hypotheses? Discounting a result simply because it relies on equilibrium conditions should only be done in the broader context of what else is known about the biology, ecology and evolutionary history of an organism.

(viii) *Having better technology makes us smarter than our predecessors*

The Discovery Channel, Wikipedia and Time Magazine are among the many sources that list the greatest scientific achievements through time. Arguably, the major advances in biology over the past few decades are technological rather than conceptual, with the major conceptual breakthroughs that set the modern framework for the many fields of biology arising primarily before this technological age. Early DNA technologies did not lend themselves easily to studies in molecular ecology. For example, when the chain-termination method of DNA sequencing was published (Sanger & Coulson 1975) the process was impractical for population research because of the vast resources needed to clone each sequence. The chemical modification and cleavage method (Maxam & Gilbert 1977) allowed direct sequencing of purified DNA, but sequencing was still technically complex and impractical for more than a decade. With the exception of a few well-funded laboratories, attempts to sequence DNA was not routine until computer and laboratory technology advanced to the point where by the early 1990s laboratories were able to easily sequence up to 100 000 base pairs if they could manage the cost (both in terms of labour and reagents). The Human Genome Project led engineers and scientists to improve the speed and accuracy of sequencing, which led to increased availability and a concordant reduction in the overall cost of sequencing (Watson 1990).

As a result of these technological advances, not just in DNA sequencing, but also in computing power and web-based manuscript review, the trend has been for more data per publication, shorter time to publication and more publications per author in the field of molecular ecology. For example, 181 recently hired tenure-track faculty world-wide had an average of 2.9 years of postdoctoral experience and an average of 11.75 (maximum 45) peer-reviewed publications at the time of hire (Marshall *et al.* 2009). By comparison, a search of Web of Knowledge (Thomson Reuters formerly ISI) most highly cited authors returned 12 that completed their doctoral dissertations before 1990 and whose CVs are available online. These researchers produced an average of only 4.6 ± 0.2 publications by three years after graduation. Likewise, a dissertation in one of the disciplines of molecular ecology prior to 1990 was typically based on sequences from a single-locus and samples sizes of

tens of individuals, whereas today dissertations are routinely expected to include several hundred lengthy sequences for multiple genes.

This increased expectation and rate of publication also results in ever more submissions to journals, which increase rejection rates due to space limitations, and that builds pressure on authors to claim the first, biggest or best study for submissions to high-impact factor journals. Claiming to be the first study to ever show some result is facilitated by an eroding knowledge of the classic literature. The awareness of the classical literature in molecular ecology is restricted by search engines that index only the past few decades of research and by the limited number of citations allowed in a publication (Pechenik *et al.* 2001; Toonen 2005). These two restrictions reduce the ability to re-discover overlooked, but important findings in the past (e.g. Wagner *et al.* 2011). All the authors of this review have reviewed papers in which disparaging remarks are made about how previous workers were misled by the limitations of the technology of their time. We must remember that these people were just like us in that they did the best they could with the technology of the day and their studies have laid the groundwork upon which our modern techniques and analyses depend. It is easy to cast stones while standing on the shoulders of giants, but we must not forget that true genius lies in the unravelling of diploid inheritance, the discovery of natural selection, the definition of an enduring species concept, the illumination of speciation, founding the field of phylogeography, or creating a journal that consolidates this field.

Discussion

In this review, we highlight some common misconceptions and oversimplifications, but the list is hardly comprehensive. Our goal is to stimulate discussion about how molecular ecologists apply their craft. Many misconceptions in the various subdisciplines of molecular ecology arise as a consequence of the huge amount of data that can be relatively easily and rapidly generated and analysed. There are many more automated DNA sequencers than classes in population genetic theory, and as self-educated molecular ecologists contribute in professional service, we sometimes see misconceptions perpetuated by journal authors, reviewers and editors.

To illustrate the growing complexity of data analysis, consider the history of computer software in population genetics. During the inception of empirical population genetic studies in the 1970s, when electrophoretic methods were first applied to population studies (Selander & Yang 1969; Utter *et al.* 1973), private programs on computer cards for mainframe computers circulated

among researchers. Knowledge of programming languages such as FORTRAN and access to uncommon and specialized equipment were necessary to implement new statistical procedures as they appeared in the literature. Today, a myriad of sophisticated computer programs take advantage of the ever-increasing capabilities of desktop computers to analyse large data sets at great speeds. Some of the misconceptions outlined in this review arise from the misapplication of these programs. Laptop computers now exceed the capabilities of the mainframe computers of 30 years ago and facilitate statistical tests based on likelihood or Bayesian methods that require millions of iterations to distinguish between models. As the field of molecular ecology rapidly grew into the current heyday, so did some of the misconceptions made along the way. Seventy years ago, Julian Huxley articulated a similar phenomenon in a heyday of organismal evolution, and coined the term 'modern synthesis' in the process.

At the end of this review, many readers will still believe that if they can properly format data for MEGA (Tamura *et al.* 2011) or ARLEQUIN (Excoffier *et al.* 2005), they do not need population genetic theory, they can pick it up along the way, or all the information they need is in the manual. Considering the high error rate (49.9%) in publications of a simple calculation of a population genetic parameter revealed by Schenekar & Weiss (2011), our answer is this: **about half of you are right**. Keeping misconceptions, inaccuracies and misstatements out of the published literature is a very complex process involving several facets. The first line of defence against the introduction of misconceptions lies with the author. It is incumbent on authors to be certain that what they are publishing is precise and accurate. This is important not only with the initial creation of the manuscript, but during the review processes as well. As disturbing as it is surprising, a survey of 179 first authors publishing in *Academy of Management Journal* and *Academy of Management Review* revealed that nearly 25% of them made manuscript changes that they thought were incorrect, in response to pressure from reviewers (Bedeian 2003). Notably these opinions are from the published (i.e. not rejected) authors. The pressure to satisfy reviewers is considerable and reinforced by the pressure to publish. As authors, we find ourselves including statements in manuscript that, though we don't believe are necessary or improve the manuscript, we think will accommodate dubious criticisms from reviewers.

The publication process also is a critical junction where misconceptions can not only become enshrined but also dispelled, and our primary defences here are the reviewers and editors who handle journal submissions (Newton 2010). As manuscript reviewers our-

selves, we routinely encounter statements that are peripheral to our expertise. We may, however, have a sense that the statement is somehow incorrect or lacking support. When this occurs, it is important for us to verify the statement. Sometimes this requires consulting the primary literature, checking citations to be sure they are appropriate and consulting with experts on that topic. This takes time but is necessary for a proper review, will help to reduce misconceptions and introduce us to new concepts along the way. Taking responsibility for and acknowledging gaps in our training is especially important because, though the number of reviewers appears to be unchanging (Vines *et al.* 2010), there is a negative correlation between the willingness of a reviewer to accept a review invitation and the reviewer's '...reviewing expertise, stature in the field, and professorial rank' (Northcraft 2001). What this means is that the people who are most qualified to catch and correct misinformation are reluctant to contribute to the review process.

As we are also associate editors who shepherd manuscripts through the review process, it is important for us to remember that the journal and the authors rely on our expertise to untangle careless, conflicting, or conflated statements *both in the manuscript and in the reviews*: to sift the intellectual wheat from the chaff (Northcraft 2001; Schwartz & Zamboanga 2009). When confronted with an unfamiliar concept, the same verification process as above needs to be conducted. As the ultimate referee, the editor should render an independent opinion as to the soundness of the research, analysis, conclusions and presentation. Equally important, however, the editor needs to review the reviews. All reviews are not equal and authors deserve an expert opinion on the veracity of criticisms and the validity of suggested changes (Tsang & Frey 2007). We sometimes hear that the peer-review process is broken: A Google search of "peer-review is broken" in December 2011 resulted in 7 450 pages (though we did not verify every page). A common theme is that editors do not take enough care with submissions (Smith 1997; Schwartz & Zamboanga 2009). Our experience supports this assertion, and the nadir of this situation is that many editors do not read the submissions. A signature of this problem is that, editors send authors the reviews along with boilerplate verbiage provided by the journal web site ("please read and respond to reviewers comments") without providing original comments. That likely is a symptom of the fast publication culture, and is fertile ground for the proliferation of misconceptions. As more scientists enter this exciting field from adjacent specialties, the publication process requires extra vigilance from all involved. Misconceptions, like deleterious mutations, should be subject to strong purifying selection.

Acknowledgements

We thank all who donate their time and effort to reviewing and editing manuscripts, and the professors who taught us theoretical population genetics, including Wyatt Anderson, Jonathan Arnold, Marjorie Asmussen, John Avise, Joseph Felsenstein, Richard Grossberg, James Hamrick, Dennis Hedgecock, Michael Turelli, Fred Utter and the faculty of the UC Davis Center for Population Biology. The inspiration and mentoring are theirs, the errors are ours. Special thanks to Fred Allendorf, Louis Bernatchez, Matt Craig, Nils Ryman, Tim Vines, Robert Vrijenhoek, Robin Waples and an anonymous review for helpful comments on the manuscript. Research funding is provided by National Science Foundation grants OCE-0627299 (SAK) and OCE-0929031 (BWB) and University of Hawaii Sea Grant Program No. NA05OAR4171048 (BWB) and the Office of National Marine Sanctuaries-HIMB partnership (MOA-2009-039/7932, SAK, BWB, RJT). This is the School of Ocean and Earth Science and Technology contribution #8561 and Hawaii Institute of Marine Science #1485.

References

- Avise JC (1994) *Molecular Markers, Natural History, and Evolution*. Chapman & Hal, Inc. New York, NY.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Baker CS, Medrano-Gonzalez L, Calambokidis J *et al.* (1998) Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. *Molecular Ecology*, **7**, 695–707.
- Ballard JW, Whitlock MC (2004) The incomplete history of mitochondria. *Molecular Ecology*, **13**, 729–744.
- Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**, 370–418.
- Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*, **312**, 570–572.
- Bedeian AG (2003) The manuscript review process: the proper roles of authors, referees, and editor. *Journal of Management Inquiry*, **12**, 331–338.
- Bermingham E, McAfferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panama Isthmus. In: *Molecular Systematics of Fishes* (eds Kocher TD, Stepien CA), pp. 113–126. Academic Press, San Diego, CA.
- Bird CE, Smouse PE, Karl SA, Toonen RJ (2011) Detecting and measuring genetic differentiation. In: *Crustacean Issues: Phylogeography and Population Genetics in Crustacea* (eds Koenemann S, Held C, Schubart C), pp. 31–55. CRC Press, Boca Raton, FL, USA.
- Birky CW Jr (2001) The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review Genetics*, **35**, 125–148.
- Birky CW Jr, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, **121**, 613–627.
- Bowen BW, Grant WS (1997) Phylogeography of the sardines (*Sardinops* spp.): assessing biogeographic models and population histories in temperate upwelling zones. *Evolution*, **51**, 1601–1610.
- Bowen BW, Bass AL, Soares L, Toonen RJ (2005) Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Molecular Ecology*, **14**, 2389–2402.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, **140**, 783–796.
- Burg TM, Gaston AJ, Winkers K, Friesen VL (2006) Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. *Molecular Ecology*, **15**, 2409–2419.
- Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus*. *Molecular Ecology*, **19**, 4339–4351.
- Carr SM, Marsall HD (2008) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the “codmother,” transatlantic vicariance and midglacial population expansion. *Genetics*, **180**, 381–389.
- Charlesworth D (2010) Don’t forget the ancestral polymorphisms. *Heredity*, **105**, 509–510.
- Coates AG, McNeill DF, Aubry M-P *et al.* (2005) An introduction to the geology of the Bocas del Toro Archipelago, Panama. *Caribbean Journal of Science*, **41**, 374–391.
- Cooper A, Lalueza-Fox C, Anderson S *et al.* (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, **409**, 704–707.
- Crandall ED, Sbrocco EJ, DeBoer TS *et al.* (2012) Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. *Molecular Biology and Evolution*, **29**, 707–719. Advance Access.
- Croizat L (1964) *Space, Time, Form: The Biological Synthesis*. Published by the author, Caracas, Venezuela.
- Cronin MA, MacNeil MD, Patton JC (2005) Variation in mitochondrial DNA and microsatellite DNA in Caribou (*Rangifer tarandus*) in North America. *Journal of Mammalogy*, **86**, 495–505.
- Crow JF, Aoki K (1984) Group selection for polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Science USA*, **81**, 6073–6077.
- Dale AI (1999) *A History of Inverse Probability from Thomas Bayes to Karl Pearson*, 2nd edn. Springer-Verlag New York, Inc. NY, New York, pp. 15.
- Daly-Engel TS, Seraphin KD, Holland KN *et al.* (2012) Global phylogeography with mixed marker analysis reveals male-mediated dispersal in the endangered scalloped hammerhead shark (*Sphyrna lewini*). *PLoS ONE*, **7**, e29986. doi:10.1371/journal.pone.0029986
- Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB (2008) Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, **17**, 4992–5007.
- Davis LA, Roalson EH, Cornell KL, McClanahan KD, Webster MS (2006) Genetic divergence and migration patterns in a North American passerine bird: implications for evolution and conservation. *Molecular Ecology*, **15**, 2141–2152.

- Dowling DK, Friberg U, Lindell J (2008) The evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology and Evolution*, **23**, 546–554.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Eble JA, Toonen RJ, Sorensen LL, Basch LV, Papastamatiou YP, Bowen BW (2011) Escaping paradise: Larval export from Hawaii in an Indo-Pacific reef fish, the Yellow Tang (*Zebrasoma flavescens*). *Marine Ecology Progress Series*, **428**, 245–258.
- Echelle AA, Hackler JC, Lack JB *et al.* (2010) Conservation genetics of the alligator snapping turtle: cytonuclear evidence of range-wide bottleneck effects and unusually pronounced geographic structure. *Conservation Genetics*, **11**, 1375–1387.
- 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.
- Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.
- Fagundes NJR, Kanitz R, Bonatto SL (2008) A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization. *PLoS ONE*, **3**, e3157.
- Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.
- Fay JC, Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial and nuclear DNA variation. *Molecular Biology and Evolution*, **16**, 1003–1005.
- Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, **23**, 691–700.
- Gaither MR, Jones SA, Sorenson L *et al.* (2011) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian Archipelago. *PLoS ONE*, **6**, e28913. doi:10.1371/journal.pone.0028913.
- Gompert Z, Fordyce JA, Forister ML, Shapiro A, Nice CC (2006) Homoploid hybrid speciation in an extreme habitat. *Science*, **314**, 1923–1925.
- Grant WS, Cheng W (2012) Incorporating deep and shallow components of genetic structure into the management of Alaskan red king crab. *Evolutionary Applications*, in press.
- Hackett SJ, Kimball RT, Reddy S *et al.* (2008) A phylogenomic study of birds reveals their evolutionary history. *Science*, **320**, 1763–1769.
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society, London, Series B*, **268**, 939–945.
- Harshman J, Braun EL, Braun MJ *et al.* (2008) Phylogenomic evidence for multiple losses of flight in ratite birds. *Proceedings of the National Academy of Sciences, USA*, **105**, 13462–13467.
- Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.
- 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.
- Hickerson MJ, Gilchrist MA, Takebayashi N (2003) Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. *Evolution*, **57**, 2216–2225.
- Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics*, **22**, 79–83.
- 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, 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.
- Hoarau G, Coyer JA, Veldsink JH, Stam WT, Olsen JL (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology*, **16**, 3606–3616.
- Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in Threespine Stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862. doi:10.1371/journal.pgen.1000862.
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nature Reviews Genetics*, **10**, 639–650.
- Houlston GJ, Olson MS (2006) Nonneutral evolution of organelle genes in *Silene vulgaris*. *Genetics*, **174**, 1983–1994.
- Hubbs CL (1943) untitled. *American Naturalist*, **77**, 365–368.
- Hudson RR (1990) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44.
- Huxley J (1943) *Evolution: The Modern Synthesis*. Harper and Brothers, New York, NY.
- Jouzel J, Masson-Delmotte V, Cattani O *et al.* (2007) Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, **317**, 793–797.
- Kaplan NL, Hudson RR, Langley CH (1989) The ‘‘hitchhiking effect’’ revised. *Genetics*, **123**, 887–899.
- Karl SA, Bowen BW, Avise JC (1992) Global population structure and male-mediated gene flow in the green turtle (*Chelonia mydas*): RFLP analysis of anonymous nuclear loci. *Genetics*, **131**, 163–173.
- Keeney DB, Heupel MR, Hueter RE *et al.* (2005) Microsatellite and mitochondrial DNA analyses of the genetic structure of blacktip shark (*Carcharhinus limbatus*) nurseries in the northwestern Atlantic, Gulf of Mexico, and Caribbean Sea. *Molecular Ecology*, **14**, 1911–1923.
- Kerth G, Mayer F, Petit E (2002) Extreme sex-biased dispersal in the communally breeding, nonmigratory Bechstein's bat (*Myotis bechsteinii*). *Molecular Ecology*, **11**, 1491–1498.
- Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513–1524.
- Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, **160**, 765–777.

- Kimball RF (1943) The great biological generalization. *The Quarterly Review of Biology*, **18**, 364–367.
- King R, Morgan BJT, Gimenez O, Brooks SP (2010) *Bayesian Analysis for Population Ecology*. CRC Press, New York, NY.
- Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.
- Kitchen A, Miyamoto MM, Mulligan CJ (2008) A three-stage colonization model for the peopling of the Americas. *PLoS ONE*, **3**, e1596.
- Lambeck K, Esat TM, Potter E-K (2002) Links between climate and sea levels for the past three million years. *Nature*, **419**, 199–206.
- Larson S, James R, Bodkin J, Staedler M, Bentzen P (2002) Microsatellite DNA and mitochondrial DNA variation in remnant and translocated sea otter (*Enhydra lutris*) populations. *Journal of Mammalogy*, **83**, 893–906.
- Larsson LC, Charlier J, Laikre L, Ryman N (2009) Statistical power for detecting divergence – organelle versus nuclear markers. *Conservation Genetics*, **10**, 1255–1264.
- Lewontin RC, Hubby JL (1966) A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, **54**, 595–605.
- Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417–1438.
- Liu M, Lu Z-C, Gao T-X, Yanagimoto T, Sakurai Y (2010) Remarkably low mtDNA control-region diversity and shallow population structure in Pacific cod *Gadus macrocephalus*. *Journal of Fish Biology*, **77**, 1071–1082.
- Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*); relevance for speciation. *Molecular Ecology*, **10**, 965–985.
- Lyrholm T, Leimar O, Johanneson B *et al.* (1999) Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations. *Proceedings of the Royal Society of London Series B*, **266**, 347–354.
- Marko PB (2002) Fossil calibration of molecular clocks and the divergence time of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution*, **19**, 2005–2021.
- Marko PB, Moran AL (2009) Out of sight, out of mind: high cryptic diversity obscures the identities and histories of geminate species in the marine bivalve subgenus *Acar*. *Journal of Biogeography*, **36**, 1861–1880.
- Marshall JC, Butters P, Callahan T *et al.* (2009) In the academic job market, will you be competitive? A case study in ecology and evolutionary biology. *Israel Journal of Ecology & Evolution*, **55**, 381–392.
- Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Science USA*, **74**, 560–564.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetic Research*, **23**, 23–35.
- McCusker MR, Bentzen P (2010) Positive relationship between genetic diversity and abundance in fishes. *Molecular Ecology*, **19**, 4852–4862.
- McLean J, Taylor E (2001) Resolution of population structure in a species with high gene flow: microsatellite variation in the eulachon (Osmeridae: *Thaleichthys pacificus*). *Marine Biology*, **139**, 411–420.
- Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.
- Mulligan CJ, Kitchen A, Miyamoto MM (2006) Comment on “population size does not influence mitochondrial genetics diversity in animals”. *Science*, **314**, 1390.
- Nakatani M, Miya M, Mabuchi K, Saitoh K, Nishida M (2011) The evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. *BMC Evolutionary Biology*, **11**, 177.
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.
- Newton DP (2010) Quality and peer review of research: an adjudicating role for editors. *Accountability in Research*, **17**, 130–145.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Northcraft H (2001) From the editors. *The Academy of Management Journal*, **44**, 1079–1080.
- Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature*, **396**, 572–575.
- O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799–1814.
- Pechenik JA, Reed JM, Russ M (2001) Should auld acquaintance be forgot: Possible influence of computer databases on citation patterns in the biological literature. *BioScience*, **51**, 583–588.
- Pérez-Losada M, Nolte MJ, Crandall KA, Shaw PW (2007) Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis*. *Molecular Ecology*, **16**, 2667–2679.
- Perry AL, Low PJ, Ellis JR *et al.* (2005) Climate change and distribution shifts in marine fishes. *Science*, **308**, 1912–1915.
- Peterson GI, Masel J (2009) Quantitative prediction of molecular clock and Ka/Ks at short timescales. *Molecular Biology and Evolution*, **26**, 2595–2603.
- Phillips MJ, Gibb GC, Crimp EA, Penny D (2010) Tinamous and Moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Systematic Biology*, **59**, 90–107.
- Piertney SB, MacColl ADC, Bacon PJ *et al.* (2000) Matrilineal genetic structure and female mediated gene flow in the red grouse (*Lagopus lagopus scoticus*): an analysis using mitochondrial DNA. *Evolution*, **54**, 279–289.
- Prager EM, Wilson AC, Osuga DT, Feeny RE (1976) Evolution of flightless land birds on southern continents—transferrin comparison shows monophyletic origin of ratites. *Journal of Molecular Evolution*, **8**, 283–294.
- Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161–165.
- Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from

- Drosophila*, mice, and humans. *Molecular Ecology and Evolution*, **13**, 735–748.
- Rato C, Carranza S, Perera A, Carretero MA, Harris DJ (2010) Conflicting patterns of nucleotide diversity between mtDNA and nDNA in the Moorish gecko, *Tarentola mauritanica*. *Molecular Phylogenetics and Evolution*, **56**, 962–971.
- Raymond M, Rousset F (1995) GENEPOP (version 3.3): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Rokas A, Atkinson RJ, Brown GS, West SA, Stone GN (2001) Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity*, **87**, 294–304.
- Roman J, Santhuff S, Moler P, Bowen BW (1999) Cryptic evolution and population structure of the alligator snapping turtle, *Macrolemys temminckii*. *Conservation Biology*, **13**, 135–142.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.
- Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.
- Ryman N, Palm S, André C *et al.* (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology*, **15**, 231–245.
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *American Journal of Human Genetics*, **67**, 718–726.
- Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, **94**, 441–448.
- Sarno RJ, Franklin WL, O'Brien SJ, Johnson WE (2001) Patterns of mtDNA and microsatellite variation in an island and mainland population of guanacos in southern Chile. *Animal Conservation*, **4**, 93–101.
- Schenecker T, Weiss S (2011) High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. *Heredity*, **107**, 511–512.
- Schlötterer C (2004) The evolution of molecular markers – just a matter of fashion. *Nature Reviews Genetics*, **5**, 63–70.
- Schmidt KP (1943) untitled. *Copeia*, **1943**, 262–263.
- Schwartz SJ, Zamboanga BL (2009) The peer-review and editorial system: ways to fix something that might be broken. *Perspectives on Psychological Science*, **4**, 54–61.
- Scribner KT, Petersen MR, Fields RL *et al.* (2001) Sex-biased gene flow in spectacled eider (Anatidae): inference from molecular markers with contrasting modes of inheritance. *Evolution*, **55**, 2105–2115.
- Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genetics*, **20**, 278–280.
- Selander RK, Yang SY (1969) Protein polymorphism and genic heterozygosity in a wild population of the house mouse (*Mus musculus*). *Genetics*, **63**, 653–667.
- Sibley CG, Ahlquist JE (1990) *Phylogeny and Classification of Birds: A Study in Molecular Evolution*. Yale University Press, New Haven CT.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.
- Smith R (1997) Peer review: reform or revolution? *BMJ*, **315**, 759.
- Stamatis C, Triantafyllidis A, Moutou KA, Mamuris M (2004) Mitochondrial DNA variation in northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. *Molecular Ecology*, **13**, 1377–1390.
- Strasser CA, Barber PH (2009) Limited genetic variation and structure in softshell clams (*Mya arenaria*) across their native and introduced range. *Conservation Genetics*, **10**, 803–814.
- Swofford DL, Selander RB (1981) BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity*, **72**, 281–283.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Theisen TC, Bowen BW, Lanier W *et al.* (2008) High connectivity on a global scale in the pelagic wahoo, *Acanthocybium solandri* (tuna family Scombridae). *Molecular Ecology*, **17**, 4233–4247.
- Toonen RJ (2005) Foundations of gregariousness in barnacles. *Journal of Experimental Biology*, **208**, 1773–1774.
- Tsang EWK, Frey BS (2007) The as-is journal review process: Let authors own their ideas. *Academy of Management Learning and Education*, **6**, 128–136.
- Turelli M, Hoffmann AA (1991) Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature*, **353**, 440–442.
- Utter FM, Allendorf FW, Hodgins HO (1973) Genetic variability and relationships in Pacific salmon and related trout based on protein variations. *Systematic Zoology*, **22**, 257–270.
- Verhoeven K, Louwye S, Eiríksson J, De Schepper S (2011) A new age model for the Pliocene-Pleistocene Tjörnes section on Iceland: Its implication for the timing of North Atlantic-Pacific palaeoceanographic pathways. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **309**, 33–52.
- Vines T, Rieseberg L, Smith H (2010) No crisis in supply of peer reviewers. *Nature*, **468**, 1041.
- Wagner D, Pochon X, Irwin L, Toonen RJ, Gates RD (2011) Azooxanthellate? Most Hawaiian black corals contain *Symbiodinium*. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1323–1328.
- 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.
- Wares JP, Barber PH, Ross-Ibarra J, Sotka EE, Toonen RJ (2006) Mitochondrial DNA and population size. *Science*, **314**, 1388–1390.
- Watson JD (1990) The human genome project: past, present & future. *Science*, **248**, 44–49.

- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 139–156.
- Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, **19**, 395–420.
-

S.A.K. focuses on the molecular ecology of marine and terrestrial organisms, with a special interest in self-evident truths. R.J.T. studies marine invertebrates but also acknowledges lessons from the chordata. W.S.G. is a gentleman of diverse genetic interests, with a special focus on marine population expansions that allegedly occurred during glacial maxima. B.W.B. studies marine vertebrates but acknowledges lessons from the other 32 phyla in the animal kingdom. All authors endorse the school of philosophy for associate editors named "Do your damn job: Read the submissions and evaluate the reviews."

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1. Methods used in simulations.

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