

Comparative phylogeographic summary statistics for testing simultaneous vicariance

M. J. HICKERSON,* G. DOLMAN† and C. MORITZ*

*Museum of Vertebrate Zoology, University of California, 3101 Valley Life Sciences Building, Berkeley, California 94720-3160, USA,

†Department of Zoology and Entomology, University of Queensland, St Lucia, 4072, Australia

Abstract

Testing for simultaneous vicariance across comparative phylogeographic data sets is a notoriously difficult problem hindered by mutational variance, the coalescent variance, and variability across pairs of sister taxa in parameters that affect genetic divergence. We simulate vicariance to characterize the behaviour of several commonly used summary statistics across a range of divergence times, and to characterize this behaviour in comparative phylogeographic datasets having multiple taxon-pairs. We found Tajima's D to be relatively uncorrelated with other summary statistics across divergence times, and using simple hypothesis testing of simultaneous vicariance given variable population sizes, we counter-intuitively found that the variance across taxon pairs in Nei and Li's net nucleotide divergence (π_{net}), a common measure of population divergence, is often inferior to using the variance in Tajima's D across taxon pairs as a test statistic to distinguish ancient simultaneous vicariance from variable vicariance histories. The opposite and more intuitive pattern is found for testing more recent simultaneous vicariance, and overall we found that depending on the timing of vicariance, one of these two test statistics can achieve high statistical power for rejecting simultaneous vicariance, given a reasonable number of intron loci (> 5 loci, 400 bp) and a range of conditions. These results suggest that components of these two composite summary statistics should be used in future simulation-based methods which can simultaneously use a pool of summary statistics to test comparative the phylogeographic hypotheses we consider here.

Keywords: coalescent, comparative phylogeography, simultaneous vicariance, statistical power, summary statistic

Received 20 March 2005; revision received 24 June 2005; accepted 25 July 2005

Introduction

One of the most central yet challenging puzzles in biogeography is whether codistributed taxon pairs share a common history of simultaneous vicariance (Darwin 1859; Nelson & Platnick 1981; Ricklefs & Schluter 1993; Avise *et al.* 1998; Johns & Avise 1998; Knowlton & Weigt 1998; Schneider *et al.* 1998; Riddle *et al.* 2000; Johnson & Cicero 2004). While conceptually simple, testing for simultaneous divergence with genetic data is obscured by differences among taxon pairs in parameters that affect genetic divergence including mutation rates, ancestral population sizes, postdivergence migration (admixture), and ancestral subdivision. Even when

there is equality in such parameters across taxon pairs, substantial variation in genetic divergence will arise from mutational and coalescent variance, especially in taxa that are heavily subdivided or have large and stable effective population sizes (Arbogast *et al.* 2002; Hickerson *et al.* 2003). In addition to these difficulties, incorrect sister species status from undetected extinction events can also hinder inference (Johnson & Cicero 2004).

Maximum-likelihood and Bayesian methods can in principle tease apart simultaneous and variable divergence histories by utilizing the full information content from the data (Bahlo & Griffiths 2000; Edwards & Beerli 2000; Nielsen & Wakeley 2001). However, extending these methods to simultaneously analyse multiple phylogeographic data sets involving complex models and idiosyncratic biogeographic histories is less tractable than simulation-based methods

Correspondence: Michael J. Hickerson, Fax: (510) 643-8238; E-mail: mhick@berkeley.edu

such as approximate likelihood and approximate Bayesian computation (ABC). Although such simulation-based approximate methods utilize less information from the data using summary statistics, they allow greater flexibility and can better handle complicated and highly parameterized models because the likelihood function, $P(\text{data} | \Phi)$ does not have to be calculated explicitly. Furthermore, simulation-based methods can easily incorporate a priori idiosyncratic biological realism or uncertainty in nuisance parameters (Pritchard *et al.* 1999; Estoup *et al.* 2001; Beaumont *et al.* 2002; Beaumont 2004). Such methods have been developed for testing demographic histories of a single taxon (Tavaré *et al.* 1997; Weiss & von Haeseler 1998; Estoup *et al.* 2004; Tallmon *et al.* 2004; Excoffier *et al.* 2005) and phylogenetic questions (Plagnol & Tavaré 2002), but histories involving sets of geographically codistributed taxa have not yet been considered.

However, such simulation-based approximate methods work best when using summary statistics that are relatively unbiased and contain relevant information regarding a parameter that is to be estimated, such as using the average

number of pairwise differences between individuals to estimate θ ($4 * \text{the effective population size} * \text{mutation rate}$) under neutrality and panmixia (Tajima 1983). When testing more complex hypotheses such as estimating the distribution of divergence times across taxon pairs, a single summary statistic might not contain enough information for robust inference across parameter space. In such cases, a simulation-based method would benefit from simultaneously using a set of summary statistics that together capture the essential features in the data.

Therefore, before embarking on a full ABC or approximate likelihood method for comparative phylogeographic parameter estimation and hypothesis testing, we must identify a pool of minimally correlated summary statistics that independently demonstrate some statistical power regarding comparative phylogeographic hypothesis testing. To this end we employ simulations to investigate the behaviour of various summary statistics under various single and multiple taxon-pair vicariance models. Specifically, we use simulations to investigate: (i) the general properties of several commonly used summary statistics (Table 1) including

Table 1 Summary statistics of single taxon-pair divergence histories. The proposed comparative phylogeographic summary statistics involve calculating variance of a subset of these summary statistics across 10 taxon pairs or covariance of pairs of a subset of these summary statistics. All summary statistics are averaged across loci within each taxon pair

Notation	Summary statistic	Reference	Expectation and variance derived under vicariance	Formula and/or description
π	Average pairwise differences	Tajima 1983	Yes	$\left(\sum_{i < j} \pi_{ij} \right) / \binom{n}{2}; \pi_{ij} = \text{differences between the } i\text{th and } j\text{th sequence; } n \text{ is the number of DNA sequences sampled}$
π_b	Average pairwise differences between populations	Takahata & Nei 1985	Yes	π restricted to between population comparisons
π_w	Average pairwise differences within populations	Takahata & Nei 1985	Yes	π restricted to within population comparisons
π_{net}	Net average pairwise differences between populations	Takahata & Nei 1985	Yes	$\pi_b - \pi_w$
S	Raw number of segregating sites	Watterson 1975	Expectation: Wakeley & Hey 1997 Variance under extreme vicariance: Hudson, Kreitman & Aguadé 1987	Number of Polymorphic sites
θ_w	Watterson's theta (number of segregating sites normalized for sample size)	Watterson 1975	No	$S / \left(\sum_{i=1}^{n-1} \frac{1}{i} \right); n \text{ is number of DNA sequences sampled}$
D	Tajima's D	Tajima 1989	No	$D = \left(\pi - S / \left(\sum_{i=1}^{n-1} \frac{1}{i} \right) \right) / \sqrt{e_1 S + e_2 S(S-1)};$ e_1 and e_2 are coefficients defined in Tajima (1989) n is number of DNA sequences sampled

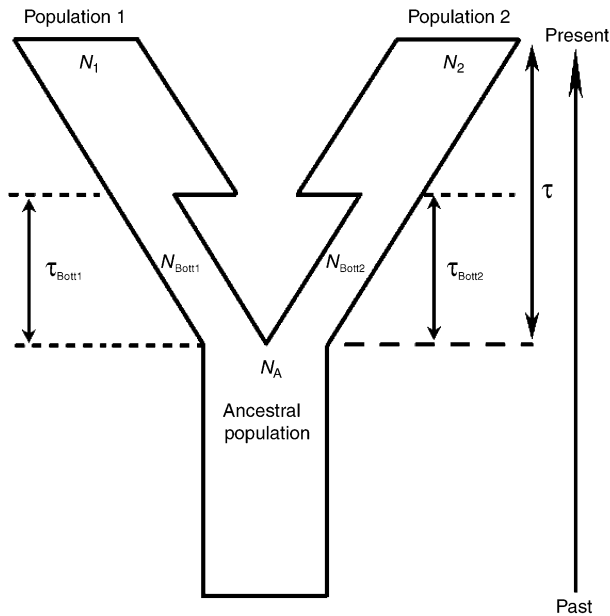


Fig. 1 A depiction of the model involving one ancestral population splitting into two-daughter populations at time τ ($4N$ generations) before the present. The population mutation parameters are θ_A , θ_1 , and θ_2 , where each θ is $4 * N * \mu$, such that μ is the per-gene per-generation DNA mutation rate and where N_A , N_1 , and N_2 are effective population sizes of the ancestral, and two-daughter populations, respectively. Under the demographic expansion model, the two-daughter populations are of sizes N_{Bott1} and N_{Bott2} , respectively, for times τ_{Bott1} and τ_{Bott2} , respectively.

their expectations, variances, and pairwise correlations throughout a range of divergence times; (ii) the variances and covariances across taxon pairs in a subset of these summary statistics (Table 1) under various simultaneous and multiple vicariance histories (Figs 1 and 2); and (iii) how statistical power to reject simultaneous vicariance improves with collecting more nuclear loci under a simple hypothesis testing framework in which the variance of a summary statistic across taxon pairs or covariance of pairs of summary statistics across taxon pairs are independently used as test statistics (Excoffier *et al.* 2000; Knowles & Maddison 2002; Hickerson & Cunningham 2005).

Results of this study will not only have general relevance to comparative phylogeographic studies, but will additionally suggest a useful pool of summary statistics to use in the full ABC framework we are concurrently developing for comparative phylogeographic studies (Beaumont *et al.* 2002).

Materials and methods

This study is partially motivated by the comparative phylogeographic data set that is emerging from the Australian Wet Tropics (AWT) in which multiple intron loci are being

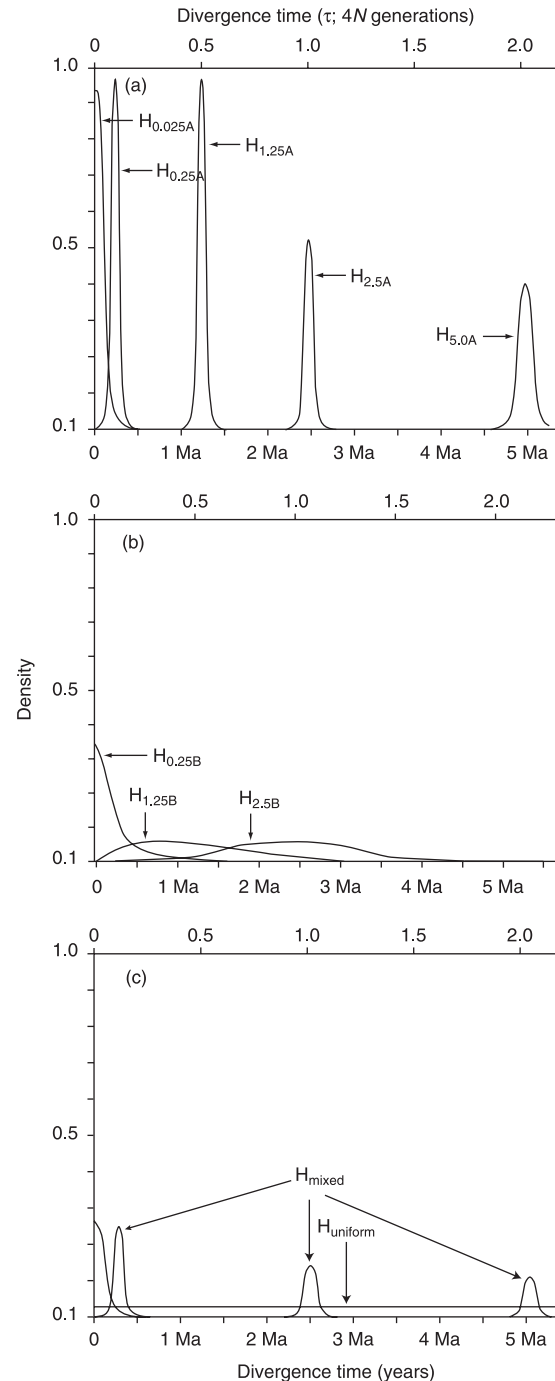


Fig. 2 Density curves depicting prior distributions that define the 10 different multiple taxon-pair vicariance hypotheses. Under each hypothesis, divergence time varies among taxon-pairs within a data set by randomly drawing from each respective prior distribution. The density curves in (a) and (b) are simulated from a gamma distribution using different means and variances for τ . The density curve depicting $H_{uniform}$ in (c) is a uniform distribution with values of τ ranging between 0 and 5.5 million years. The density curve depicting H_{mixed} in (c) is a mixture of gamma distributions $H_{0.025A}$, $H_{0.25A}$, $H_{2.5A}$, and $H_{5.0A}$ with equal probability. Divergence time is alternatively given in units of $4N$ generations and years.

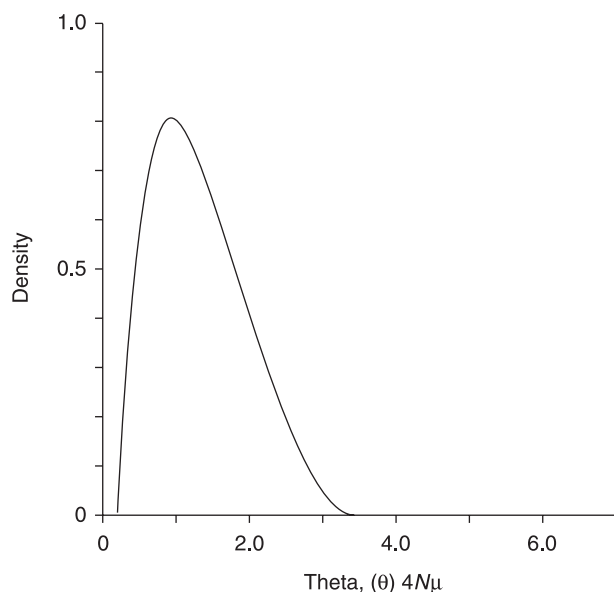


Fig. 3 Density curve depicting the prior distribution defining among-taxon-pair variation in θ ($4N\mu$) used in the 10 taxon-pair vicariance simulations.

collected across taxon pairs that are codistributed across a putative historical barrier to gene flow (the Black Mountain barrier, Schneider *et al.* 1998). Therefore, we incorporate a priori information on effective population sizes (Fig. 3), mutation rates, generation times and divergence times that are based on estimates from seven introns collected from two *carlia* skink taxon pairs (Dolman & Phillips 2004; Dolman & Moritz in review). The simulated single taxon-pair data sets consisted of samples of 12 diploid individuals per sister taxon and 400 bp introns per locus (Fig. 1). The simulated comparative phylogeographic data sets were identical to this, yet consisted of 10 taxon pairs per data set.

Summary statistics

We chose to explore summary statistics based on whether they were likely to contain information relevant to the parameters in our model, such as effective population sizes, divergence times, and other demographic processes (Table 1). With this in mind we initially chose two compound summary statistics (π_{net} and Tajima's D) as well as their components (π_w and π_b from the former and S , π , and θ_w from the latter). Although Tajima's D , S , π , and θ_w are usually quantities that are calculated from a single species or population, here we calculate each of these summary statistics once per taxon pair as if they were being collected from a single population or species. On the other hand, π_{net} and its components (π_w and π_b) are calculated with each sister taxon treated as a distinct population, such that π_b is the average pairwise differences between a sister pair, π_w

is the average pairwise differences within the two sister taxa, and $\pi_{\text{net}} = \pi_b - \pi_w$ (Table 1). With respect to divergence time, it is known that π_b , π_{net} , Tajima's D , S , π , θ_w all increase as divergence time increases (Takahata & Nei 1985; Hudson *et al.* 1987; Tajima 1989; Simonsen *et al.* 1995). With respect to population sizes, we can expect π_w , S , π , and θ_w to all contain information (Watterson 1975; Tajima 1983; Fu 1995; Vasco *et al.* 2001), and we should expect Tajima's D to extract information regarding population expansion (Simonsen *et al.* 1995).

There were some commonly used summary statistics that we did not consider because they are known to contain less information for our parameters of interest. For example, we considered π_{net} instead of F_{ST} , a commonly used metric in population genetic studies. Although these two metrics are very similar in that they both measure the ratio of interpopulation and intrapopulation genetic variance, we initially chose the former because it is not upwardly bounded with greater divergence times, while the latter has a maximum value of 1.0. Likewise we did not choose to explore haplotype diversity, the number of haplotypes or the number of singletons, because these measures do not have strong bearing on divergence times. We also did not explore Fu and Li's D or F statistic because these measures are known to behave very similar to Tajima's D (Fu & Li 1993), but are only more powerful than Tajima's D when an outgroup is known (Simonsen *et al.* 1995), a condition we do not assume.

Expectation, variance and correlation of summary statistics

To select candidate summary statistics from our initial pool (Table 1), we determined their mean and variance throughout a range of divergence times and determined the correlations between all pairs of these seven summary statistics under the various divergence times. The three criteria for selecting candidate summary statistics for use in the multiple taxon-pair analyses were: (i) a strong linear relationship between divergence time and the mean of the summary statistic; (ii) low variance of the summary statistic; and (iii) low correlation between a summary statistic and other summary statistics satisfying 1 and 2 (through a range of τ). When a pair of summary statistics was shown to be heavily correlated, elimination was based on the degree to which criteria 1 and 2 were satisfied by either summary statistic and whether either of the two summary statistics were correlated with other summary statistics.

The mean and variance of the seven summary statistics (Table 1) were calculated from 100 000 single taxon-pair data sets simulated under 10 divergence times discretely ranging from 0.0 to 2.0 coalescent time units (i.e. zero to approximately 5 million years ago (Ma) when scaled by $4N$ generations, generation time of 3 years and $\theta = 1.0$). The 3-year generation time was taken from the two previously

Table 2 Parameters used for simulation of vicariance histories depicted in Fig. 1

Parameter	Description	Values under single taxon-pair vicariance histories	Prior distributions under 10 taxon-pair vicariance histories
τ	Divergence time before the present in units of $4N$ generations	9 fixed values discretely ranging between 0.0–2.0 ($4N$ generations)	Gamma distributed under 11 different hypotheses defined by various values of $E(\tau)$ and $\text{var}(\tau)$ shown in Fig. 2a, b. Uniformly distributed under one hypothesis with τ between 0.0 and 5.0 $4N$ generations (Fig. 2c), and a mixed gamma distribution under one hypothesis with four means (Fig. 2c)
θ ($4 * N * \mu$)	$4 * \text{effective population size} * \text{DNA mutation rate}$ with $\alpha = 3.50$ and $\beta = 0.288$ (Fig. 3)	1.0	Gamma distributed priors (per locus per generation); $N_A = N_1 = N_2$ where N_1 and N_2 correspond to daughter taxon 1 and 2, respectively, and N_A corresponds to the ancestral population (Fig. 1)
N_{Bott1} and N_{Bott2}	Effective population size of daughter taxa 1 and 2 during bottleneck (Fig. 1)	No bottleneck or fixed at $0.5N_1$ and $0.5N_2$	No bottleneck or fixed at $0.5N_1$ and $0.5N_2$
τ_{Bott1} and τ_{Bott2}	Longevity of bottleneck subsequent to τ in daughter taxon 1 and 2	No bottleneck or fixed at 0.5τ	No bottleneck or fixed at 0.5τ

mentioned skink taxon pairs (Dolman & Moritz in review). The population mutation parameter, θ ($4N\mu$), was fixed at 1.0, such that μ , the mutation rate was 1.2×10^{-6} per locus per generation following an infinite sites model. Simulations were conducted under a model without demographic expansion and a model that included a bottleneck after divergence such that the demographic expansion parameters were $N_{\text{Bott1}} = 0.5N_1$, $N_{\text{Bott2}} = 0.5N_1$, $\tau_{\text{Bott1}} = 0.5\tau$, and $\tau_{\text{Bott2}} = 0.5\tau$ (Fig. 1; Table 2).

Although it is preferable to derive the expectation and variance of these summary statistics rather than use simulations, the range of models and summary statistics we explore makes this endeavour beyond the scope of this study. In cases where the first two moments have been previously derived under vicariance models, we compare the simulated expectations and variances to the derived values as a check.

Multiple taxon-pair divergence model

For the multiple taxon-pair simulations, we implement a hierarchy of parameters. The 'higher level' parameters describe the mean and variance in divergence time (τ) and θ within each simulated multiple taxon-pair data set, while the 'lower level' parameters are the values of τ and θ for each taxon pair within each simulated data set.

Each simulated data set consisted of 10 taxon pairs that diverged at time τ , and samples from each sister taxon contained 12 diploid individuals (Fig. 1). For each of the 10 000 data sets simulated under each of the 10 divergence time hypotheses (Fig. 2), we calculated several comparative

phylogeographic summary statistics (Table 1). For eight out of 10 of these divergence time hypotheses, τ varied across taxon pairs by drawing it from a particular gamma distribution that was constrained by a particular mean and variance of τ (Fig. 2a, b). For the two other divergence time hypotheses, τ was either drawn from a uniform prior distribution or a mixed gamma distribution with four peaks corresponding to major events of climate change (Fig. 2c): the last glacial maximum (20 000 years BP); the previous glacial maximum (150 000 years BP); the Pliocene–Pleistocene transition (2 million years ago (Ma)); and the Pliocene–Miocene boundary (5 Ma).

For each simulated taxon pair, the population mutation parameter (θ) of the ancestral (θ_A) and two daughter populations (θ_1 and θ_2 at $\tau = 0$) were held equal to each other while θ_1 and θ_2 were equally rescaled during the bottleneck phase (Fig. 1). These parameters (θ_A , θ_1 and θ_2) varied among taxon pairs within each 10 taxon-pair data set according to a gamma distribution with $\alpha = 3.5$ and $\beta = 0.288$, approximately corresponding to values of θ that range between 0.0 and 3.0 (mean of 1.02 and SD = 0.55; Fig. 3). This gamma distribution describes the range of θ that is normally found for intron loci and the mean and variance were specifically estimated from seven intron loci collected from two taxon pairs of *carlia* skinks (Dolman & Phillips 2004; Dolman & Moritz in review) using the program IM (Hey & Nielsen 2004). We explored multiple taxon-pair divergence models with and without a bottleneck. In the former case, the demographic expansion parameters were fixed at $N_{\text{Bott1}} = 0.5N_1$, $N_{\text{Bott2}} = 0.5N_1$, $\tau_{\text{Bott1}} = 0.5\tau$, and $\tau_{\text{Bott2}} = 0.5\tau$ (Fig. 1; Table 2).

Simulation procedure

The overall framework for simulating the single and 10 taxon-pair data sets according to a coalescent model used standard methodology (Hudson 1983, 1990) by drawing parameters used for each simulated iteration from their respective prior distributions and subsequently calculating the distributions of the summary statistics over a set number of iterations. Specifically, a C program drew the parameters from their respective prior distributions, and these parameter values were subsequently used as inputs for Hudson's coalescent simulator (Hudson 2002). Subsequently, another C program generated summary statistics for each simulated sample, constructed their distributions and statistical powers (a modified version of Hudson's sample_stats, 2003). These three C programs worked in consort by using a Perl script. The following general procedure was used to generate the summary statistics under each multiple taxon-pair divergence hypothesis:

- Step 1. Randomly draw N_{Bott1} , N_{Bott2} , τ_{Bott1} and τ_{Bott2} for each data set from corresponding prior distributions or hold at fixed value depending on model.
- Step 2. Randomly draw θ for each taxon pair from a gamma distribution (Fig. 3) and draw τ for each taxon pair from the prior distribution depicting a particular vicariance hypothesis (Fig. 2).
- Step 3. Using these randomly drawn parameter values, simulate the independent genealogical histories with mutations for each locus (400 bp) within each taxon pair as described in Hudson (1990).
- Step 4. Repeat steps two and three for each taxon pair within a single data set.
- Step 5. Compute summary statistics (Table 1) for each 10 taxon-pair data set.
- Step 6. Repeat steps one through five 10 000 times for each divergence time hypothesis (Fig. 2).

Statistical power

We report statistical power for the summary statistics that were found to best discriminate among near-simultaneous and variable τ hypotheses. Specifically, we compared the summary statistics and calculated their respective statistical powers among pairs of hypotheses. In each comparison, we compared a near simultaneous τ hypothesis and a variable τ hypothesis, and always considered the former to be the null hypotheses. Each null hypothesis, H_0 ($H_{0.025A}$, $H_{0.25A}$, $H_{1.25A}$, $H_{2.5A}$ and $H_{5.0A}$), was compared to a respective

alternative hypothesis, H_{Alt} ($H_{0.25B}$, $H_{0.25B'}$, $H_{1.25B}$, $H_{2.5B}$, H_{mixed} and H_{uniform} ; Fig. 2). Statistical power was defined as the probability of rejecting the null hypothesis given that the data were generated by the alternative hypothesis and an α probability of falsely rejecting this null hypothesis. The empirical cut-off points for calculating the power of each summary statistic were obtained by simulating each summary statistic under the null hypothesis. Because the distribution of each summary statistic simulated under the null hypothesis was usually bounded by zero, we chose our empirical cut-off points to correspond to the upper 5% significance level (one-tailed test). For each test, 10 000 simulations of the data under null hypothesis were compared to 10 000 simulations of the data under the alternative hypothesis.

Results

Expectation, variance and correlation of summary statistics

Using among-taxon-pair variance in a summary statistic to test for variable divergence among taxon pairs can be successful if the summary statistic is correlated with τ (criterion 1). Given the models with and without a bottleneck, all of the summary statistics tended to increase with increasing τ , with the exception of π_w and Tajima's D . The former remained close to 1.0 and makes it a less suitable summary statistic for testing multiple taxon-pair hypotheses (criterion 1). Although Tajima's D initially decreases with increasing τ , it eventually increases with increasing τ (criterion 1; Fig. 4). Considering criterion 2, variance in four of the seven summary statistics (π , π_b , π_{net} , S) greatly increased with increasing divergence time under both demographic models (criterion 2; Fig. 4). We chose to only eliminate S , given that its increase in variance was most severe and that the others remained useful by strongly satisfying criterion 1.

The correlations between some pairs of summary statistics were often heavily dependent on τ , while the correlation of other pairs remained independent of τ (Fig. 5). Among summary statistics satisfying criteria 1 and 2, π and π_b were heavily correlated at increasing τ , and both became heavily correlated with π_{net} with increasing τ (Fig. 5). Because π_b was more strongly correlated with π_{net} , π_b was eliminated from the subsequent multiple taxon-pair analysis. S , the number of segregating sites was heavily correlated with π , π_b , π_{net} and θ_w . On the other hand, Tajima's D , π and θ_w were never heavily correlated with each other under either demographic model, and while all three became correlated with π_{net} with older τ , this was to a lesser degree than correlations involving π_b or S (Fig. 5). Overall, Tajima's D was the least correlated with any of the other summary statistics.

Based on the three criteria, we selected π , π_{net} , D , and θ_w to explore in comparative phylogeographic models. The

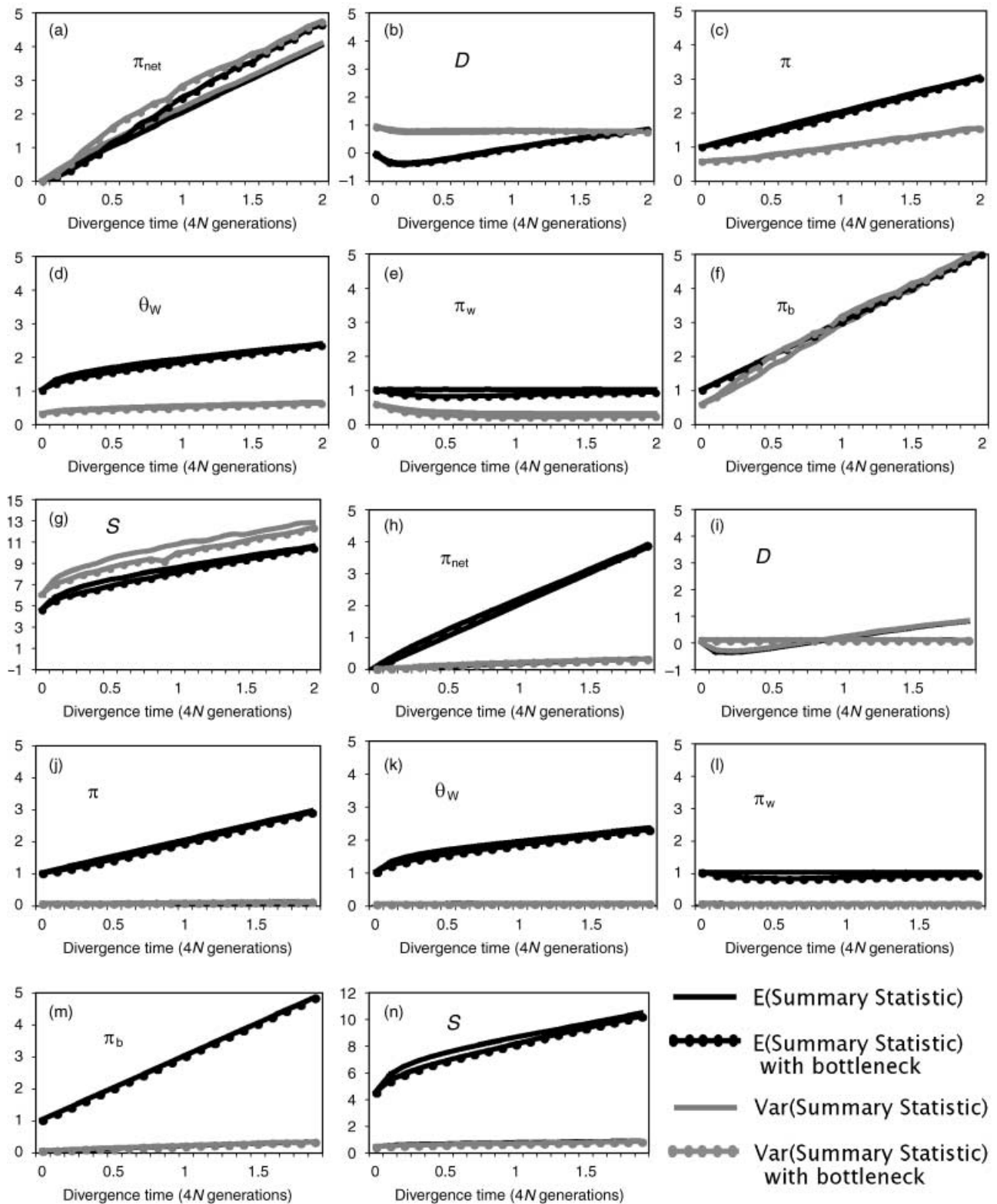


Fig. 4 The expectation and variance of the seven summary statistics (Table 1) calculated from 100 000 single taxon-pair data sets simulated under a range of divergence times. Single-locus data sets are considered in (a) through (g), and 15 loci data sets are depicted in (h) through (n). In the later case, the summary statistics are averaged across loci.

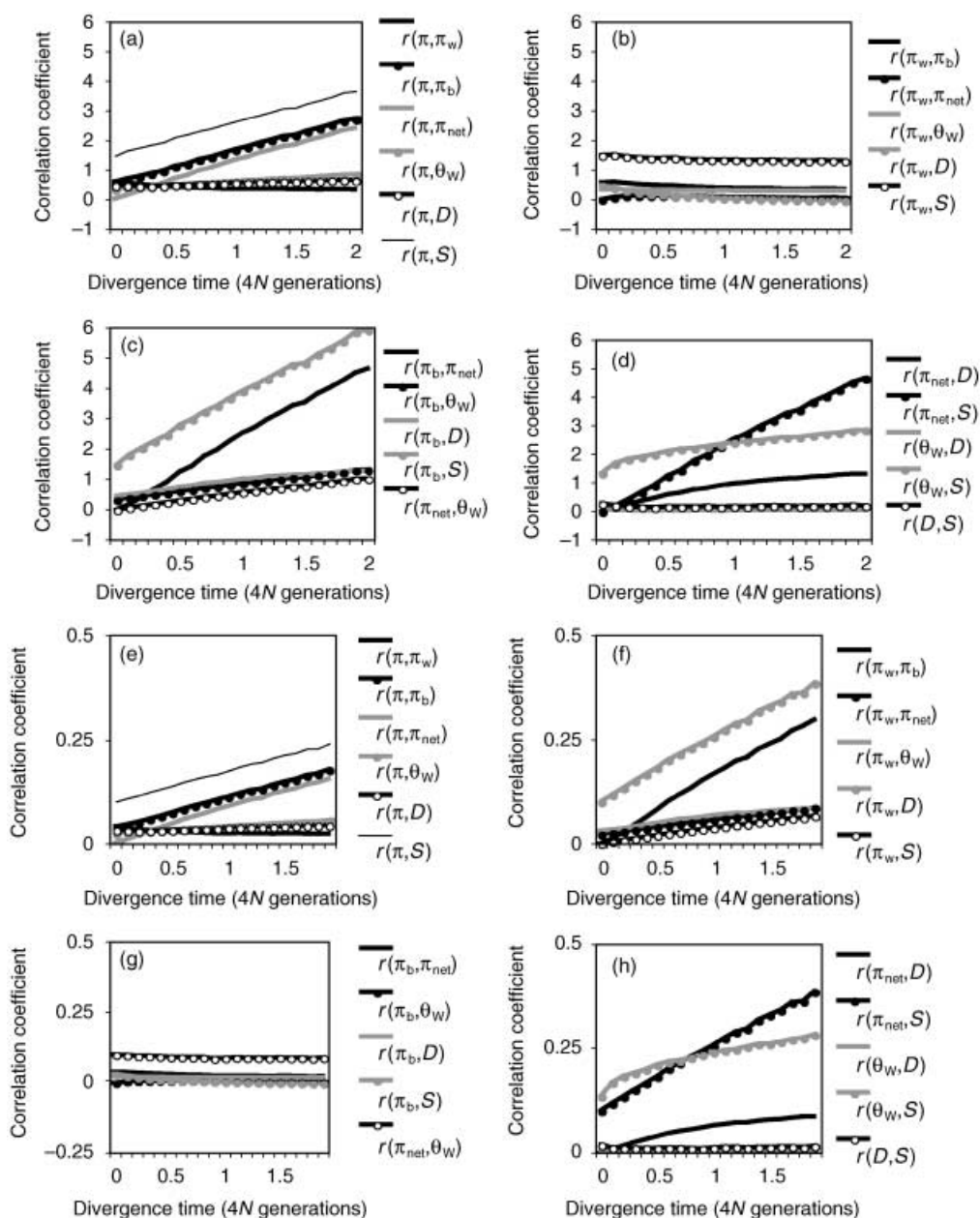


Fig. 5 Correlation coefficients among pairs of summary statistics across a range of τ (0–2.0; $4N$ generations). Values from each τ are from 100 000 single taxon-pair simulation replicates under a model. Single-locus values are (a) through (d) and the 15 loci values are depicted in (e) through (h).

results of the simulations were consistent with previous simulation studies (Simonsen *et al.* 1995), and previous derivations of the means and variances for both two-population vicariance and panmictic ($\tau = 0$) models (Takahata & Nei 1985; Hudson *et al.* 1987; Fu 1995; Wakeley & Hey 1997; Vasco *et al.* 2001).

Ten taxon-pair divergence model

Of the 10 comparative phylogeographic summary statistics we explored (Table 1), six achieved high statistical power

(> 0.5) in testing near-simultaneous divergence given modest numbers of loci (Table 3). The effect of a demographic expansion involving a bottleneck subsequent to divergence only had a slight negative or positive affect on the power of the summary statistics (Table 3b).

With and without a bottleneck, $\text{var}(\pi_{net})$, $\text{var}(D)$, and $\text{cov}(\pi_{net}, D)$ were the most useful summary statistics, but their power depended on which hypothesis was being tested. Specifically, $\text{var}(\pi_{net})$ was superior to the other summary statistics when testing for recent simultaneous vicariance ($H_{0.025A}$ vs. $H_{0.25B}$; $H_{0.25A}$ vs. $H_{0.25B}$). Conversely,

Table 3 Statistical power of comparative phylogeographic summary statistics to reject H_0 given H_{Alt} is the true vicariance hypothesis. The considered vicariance hypotheses are shown on Fig. 1. Statistical power was calculated for 1, 5, 15 and 30 loci, and under a model without an instantaneous bottleneck (a), and with a 50% reduction in population size during τ_{Bott1} and τ_{Bott2}

(a)						(b)					
H_0	1 locus	5 loci	15 loci	30 loci	H_{Alt}	H_0	1 locus	5 loci	15 loci	30 loci	H_{Alt}
$\text{var}(\pi_{\text{net}})$						$\text{var}(\pi_{\text{net}})$					
$H_{0.025A}$	0.90	0.95	0.98	0.99	$H_{0.25B}$	$H_{0.025A}$	0.87	0.95	0.97	0.98	$H_{0.25B}$
$H_{0.25A}$	0.35	0.41	0.59	0.71	$H_{0.25B}$	$H_{0.25A}$	0.23	0.34	0.60	0.67	$H_{0.25B}$
$H_{1.25A}$	0.27	0.39	0.48	0.54	$H_{1.25B}$	$H_{1.25A}$	0.21	0.26	0.39	0.43	$H_{1.25B}$
$H_{2.5A}$	0.16	0.20	0.21	0.28	$H_{2.5B}$	$H_{2.5A}$	0.12	0.18	0.23	0.27	$H_{2.5B}$
$H_{5.0A}$	0.10	0.13	0.12	0.10	H_{mixed}	$H_{5.0A}$	0.10	0.11	0.10	0.14	H_{mixed}
$H_{5.0A}$	0.10	0.12	0.11	0.13	H_{uniform}	$H_{5.0A}$	0.09	0.12	0.15	0.17	H_{uniform}
$\text{var}(D)$						$\text{var}(D)$					
$H_{0.025A}$	0.06	0.08	0.10	0.11	$H_{0.25B}$	$H_{0.025A}$	0.07	0.10	0.14	0.16	$H_{0.25B}$
$H_{0.25A}$	0.12	0.16	0.23	0.28	$H_{0.25B}$	$H_{0.25A}$	0.13	0.13	0.21	0.22	$H_{0.25B}$
$H_{1.25A}$	0.13	0.33	0.71	0.88	$H_{1.25B}$	$H_{1.25A}$	0.17	0.25	0.55	0.77	$H_{1.25B}$
$H_{2.5A}$	0.41	0.66	0.70	0.95	$H_{2.5B}$	$H_{2.5A}$	0.18	0.25	0.60	0.90	$H_{2.5B}$
$H_{5.0A}$	0.42	0.72	1.00	1.00	H_{mixed}	$H_{5.0A}$	0.29	0.61	0.98	1.00	H_{mixed}
$H_{5.0A}$	0.47	0.78	1.00	1.00	H_{mixed}	$H_{5.0A}$	0.34	0.67	1.00	1.00	H_{uniform}
$\text{cov}(\pi_{\text{net}}, D)$						$\text{cov}(\pi_{\text{net}}, D)$					
$H_{0.025A}$	0.06	0.39	0.18	0.09	$H_{0.25B}$	$H_{0.025A}$	0.05	0.07	0.06	0.10	$H_{0.25B}$
$H_{0.25A}$	0.05	0.10	0.05	0.05	$H_{0.25B}$	$H_{0.25A}$	0.06	0.09	0.11	0.04	$H_{0.25B}$
$H_{1.25A}$	0.29	0.44	0.75	0.92	$H_{1.25B}$	$H_{1.25A}$	0.25	0.29	0.34	0.78	$H_{1.25B}$
$H_{2.5A}$	0.24	0.57	0.51	0.68	$H_{2.5B}$	$H_{2.5A}$	0.24	0.26	0.75	0.62	$H_{2.5B}$
$H_{5.0A}$	0.35	0.61	0.82	0.88	H_{mixed}	$H_{5.0A}$	0.36	0.42	0.76	0.87	H_{mixed}
$H_{5.0A}$	0.39	0.64	0.85	0.93	H_{uniform}	$H_{5.0A}$	0.41	0.48	0.82	0.92	H_{uniform}
$\text{cov}(\pi_{\text{net}}, \theta_W)$						$\text{cov}(\pi_{\text{net}}, \theta_W)$					
$H_{0.025A}$	0.06	0.73	0.75	0.80	$H_{0.25B}$	$H_{0.025A}$	0.56	0.78	0.78	0.77	$H_{0.25B}$
$H_{0.25A}$	0.05	0.24	0.26	0.25	$H_{0.25B}$	$H_{0.25A}$	0.11	0.19	0.26	0.23	$H_{0.25B}$
$H_{1.25A}$	0.21	0.20	0.20	0.22	$H_{1.25B}$	$H_{1.25A}$	0.12	0.17	0.22	0.20	$H_{1.25B}$
$H_{2.5A}$	0.21	0.20	0.19	0.18	$H_{2.5B}$	$H_{2.5A}$	0.08	0.13	0.14	0.17	$H_{2.5B}$
$H_{5.0A}$	0.05	0.10	0.08	0.06	H_{mixed}	$H_{5.0A}$	0.05	0.07	0.06	0.08	H_{mixed}
$H_{5.0A}$	0.06	0.08	0.09	0.11	H_{uniform}	$H_{5.0A}$	0.06	0.06	0.08	0.11	H_{uniform}
$\text{var}(\pi)$						$\text{var}(\pi)$					
$H_{0.025A}$	0.06	0.16	0.18	0.18	$H_{0.25B}$	$H_{0.025A}$	0.08	0.13	0.15	0.17	$H_{0.25B}$
$H_{0.25A}$	0.05	0.09	0.14	0.11	$H_{0.25B}$	$H_{0.25A}$	0.09	0.12	0.11	0.11	$H_{0.25B}$
$H_{1.25A}$	0.09	0.16	0.16	0.18	$H_{1.25B}$	$H_{1.25A}$	0.11	0.16	0.18	0.18	$H_{1.25B}$
$H_{2.5A}$	0.10	0.13	0.12	0.18	$H_{2.5B}$	$H_{2.5A}$	0.12	0.15	0.15	0.18	$H_{2.5B}$
$H_{5.0A}$	0.05	0.08	0.06	0.05	H_{mixed}	$H_{5.0A}$	0.05	0.06	0.05	0.07	H_{mixed}
$H_{5.0A}$	0.06	0.07	0.07	0.09	H_{uniform}	$H_{5.0A}$	0.05	0.07	0.08	0.11	H_{uniform}
$\text{var}(\theta_W)$						$\text{var}(\theta_W)$					
$H_{0.025A}$	0.06	0.30	0.30	0.33	$H_{0.25B}$	$H_{0.025A}$	0.16	0.21	0.23	0.24	$H_{0.25B}$
$H_{0.25A}$	0.05	0.12	0.11	0.10	$H_{0.25B}$	$H_{0.25A}$	0.08	0.10	0.10	0.10	$H_{0.25B}$
$H_{1.25A}$	0.09	0.12	0.10	0.13	$H_{1.25B}$	$H_{1.25A}$	0.12	0.13	0.12	0.12	$H_{1.25B}$
$H_{2.5A}$	0.04	0.06	0.08	0.12	$H_{2.5B}$	$H_{2.5A}$	0.10	0.12	0.11	0.15	$H_{2.5B}$
$H_{5.0A}$	0.05	0.05	0.06	0.03	H_{mixed}	$H_{5.0A}$	0.05	0.06	0.04	0.06	H_{mixed}
$H_{5.0A}$	0.06	0.05	0.05	0.07	H_{uniform}	$H_{5.0A}$	0.04	0.06	0.06	0.07	H_{uniform}
$\text{cov}(\pi, \pi_{\text{net}})$						$\text{cov}(\pi, \pi_{\text{net}})$					
$H_{0.025A}$	0.06	0.80	0.73	0.79	$H_{0.25B}$	$H_{0.025A}$	0.61	0.79	0.77	0.76	$H_{0.25B}$
$H_{0.25A}$	0.05	0.21	0.23	0.24	$H_{0.25B}$	$H_{0.25A}$	0.13	0.16	0.22	0.21	$H_{0.25B}$
$H_{1.25A}$	0.29	0.26	0.28	0.31	$H_{1.25B}$	$H_{1.25A}$	0.16	0.22	0.27	0.26	$H_{1.25B}$
$H_{2.5A}$	0.04	0.07	0.15	0.22	$H_{2.5B}$	$H_{2.5A}$	0.07	0.16	0.19	0.21	$H_{2.5B}$
$H_{5.0A}$	0.05	0.08	0.09	0.06	H_{mixed}	$H_{5.0A}$	0.06	0.07	0.07	0.09	H_{mixed}
$H_{5.0A}$	0.05	0.07	0.08	0.12	H_{uniform}	$H_{5.0A}$	0.06	0.06	0.11	0.13	H_{uniform}

Table 3 Continued

(a)						(b)					
H ₀	1 locus	5 loci	15 loci	30 loci	H _{Alt}	H ₀	1 locus	5 loci	15 loci	30 loci	H _{Alt}
cov(π , D)						cov(π , D)					
H _{0.025A}	0.06	0.05	0.04	0.04	H _{0.25B}	H _{0.025A}	0.06	0.08	0.07	0.05	H _{0.25B}
H _{0.25A}	0.05	0.10	0.14	0.14	H _{0.25B}	H _{0.25A}	0.05	0.12	0.14	0.16	H _{0.25B}
H _{1.25A}	0.29	0.29	0.45	0.55	H _{1.25B}	H _{1.25A}	0.16	0.23	0.43	0.51	H _{1.25B}
H _{2.5A}	0.24	0.32	0.37	0.46	H _{2.5B}	H _{2.5A}	0.15	0.25	0.38	0.49	H _{2.5B}
H _{5.0A}	0.35	0.53	0.58	0.61	H _{mixed}	H _{5.0A}	0.27	0.34	0.55	0.65	H _{mixed}
H _{5.0A}	0.39	0.58	0.64	0.72	H _{uniform}	H _{5.0A}	0.28	0.37	0.61	0.76	H _{uniform}
cov(D , θ_W)						cov(D , θ_W)					
H _{0.025A}	0.06	0.06	0.04	0.03	H _{0.25B}	H _{0.025A}	0.06	0.07	0.06	0.05	H _{0.25B}
H _{0.25A}	0.05	0.09	0.09	0.10	H _{0.25B}	H _{0.25A}	0.08	0.10	0.11	0.12	H _{0.25B}
H _{1.25A}	0.29	0.23	0.13	0.41	H _{1.25B}	H _{1.25A}	0.13	0.20	0.34	0.39	H _{1.25B}
H _{2.5A}	0.24	0.23	0.24	0.37	H _{2.5B}	H _{2.5A}	0.16	0.23	0.29	0.36	H _{2.5B}
H _{5.0A}	0.35	0.42	0.45	0.48	H _{mixed}	H _{5.0A}	0.22	0.34	0.46	0.53	H _{mixed}
H _{5.0A}	0.37	0.46	0.49	0.55	H _{uniform}	H _{5.0A}	0.20	0.36	0.49	0.57	H _{uniform}
cov(π , θ_W)						cov(π , θ_W)					
H _{0.025A}	0.06	0.26	0.23	0.25	H _{0.25B}	H _{0.025A}	0.11	0.16	0.18	0.20	H _{0.25B}
H _{0.25A}	0.05	0.09	0.12	0.10	H _{0.25B}	H _{0.25A}	0.09	0.11	0.11	0.11	H _{0.25B}
H _{1.25A}	0.09	0.13	0.13	0.16	H _{1.25B}	H _{1.25A}	0.09	0.15	0.03	0.13	H _{1.25B}
H _{2.5A}	0.04	0.04	0.09	0.15	H _{2.5B}	H _{2.5A}	0.06	0.12	0.14	0.16	H _{2.5B}
H _{5.0A}	0.05	0.08	0.06	0.04	H _{mixed}	H _{5.0A}	0.07	0.06	0.04	0.07	H _{mixed}
H _{5.0A}	0.05	0.08	0.08	0.10	H _{uniform}	H _{5.0A}	0.06	0.06	0.05	0.09	H _{uniform}

the power of var(D) was superior when testing more ancient simultaneous vicariance hypotheses (H_{2.5A} vs. H_{2.5B}; H_{5.0A} vs. H_{mixed}; H_{5.0A} vs. H_{uniform}; Table 3), and cov(π , D) achieved the highest power when testing hypotheses involving moderate divergence times (H_{1.25A} vs. H_{1.25B}; Table 3).

Discussion

Comparative phylogeographic inference

Inferring the temporal distribution of vicariance across codistributed taxon pairs has been a vexing problem for biogeographers (Darwin 1859; MacArthur & Wilson 1967; Case 1983; Brown 1995; Avise 2000). While molecular phylogeography provides hope in this endeavour (Avise 2000), two problematic issues that are informed from theoretical population genetics illustrate the intrinsic difficulty in solving this riddle of comparative phylogeography (Edwards & Beerli 2000; Arbogast *et al.* 2002).

The first problematic issue is that intrinsic coalescent and mutation variance will result in variable genetic divergences across taxon pairs under most situations (Arbogast *et al.* 2002), even if taxon pairs are identical in all characteristics. This is expected to be found when using a single locus such as mitochondrial DNA because each mitochondrial tree

is but a single realization of a highly variable process, such that biased inferences easily arise without statistically incorporating this inherent variance (Nichols 2001; Knowles & Maddison 2002; Hudson & Turelli 2003). This problematic issue is most clearly illustrated in the 'natural experiment' of the Panamanian Isthmus splitting marine taxa into sister taxon pairs approximately 3.1 Ma (Jordan 1908; Coates *et al.* 1992). The variation in mitochondrial divergences across sister taxon pairs might first appear to signify variation in divergence times (Lessios *et al.* 2001; Marko 2002), yet such patterns are expected under both simultaneous and variable vicariance histories due to these two sources of intrinsic variance that become most apparent in single-locus data sets (Hickerson *et al.* 2003).

Despite the steady stream of cautionary reminders about the hazards of basing phylogeographic inference on a single locus (Takahata 1989; Moore 1995; Hoelzer 1997; Maddison 1997; Bermingham & Moritz 1998; Kuhner *et al.* 1998; Edwards & Beerli 2000; Beerli & Felsenstein 2001; Hare & Palumbi 2001; Hudson & Turelli 2003; Ballard & Whitlock 2004), the high cost of developing nuclear intron markers in a broad range of 'nonmodel' taxa have so far prevented widespread use of introns or other independently segregating sets of linked SNPs. Assuming that future advances will allow nuclear loci to be more routinely collected, there should be guidance in how many intron loci are sufficient

for sound inference across a range of phylogeographic questions (Edwards & Beerli 2000; Hare & Palumbi 2001; Zhang & Hewitt 2003; Aitken *et al.* 2004), and the results of this study demonstrate that previously intractable problems in phylogeography can be solved with reasonable numbers of loci.

The second problematic issue arises from among-taxon-pairs variability in the parameters that affect genetic divergences other than the actual timings of vicariance. This can include differences in among-taxon-pair mutation rates, ancestral population sizes, postdivergence migration rates (admixture), degrees of subdivision, and/or bottleneck magnitudes. Again, the Panama Isthmus example illustrates this difficulty in that the observed variability in single-locus genetic divergences across taxon pairs could have arisen from among-taxa differences in any of these confounding parameters. This second issue is daunting, but there is hope in tackling it by identifying the simplest underlying population genetic model based on a priori biological and/or nonbiological hypotheses (Wakeley 2004), and incorporating among-taxa variability in these parameters directly into the analysis via a prior distribution. It has been suggested that ABC approaches are well suited in this endeavour (Beaumont & Rannala 2004). Our primary objective here is to identify a pool of summary statistics that can extract the important features of phylogeographic data sets that are relevant to testing comparative phylogeographic hypotheses within a future simulation-based framework such as ABC.

Summary statistics for comparative phylogeographic inference

Our initial single taxon-pair simulations identify π , π_{net} , θ_W , and Tajima's D to be a potentially useful set of minimally correlated summary statistics that contain useful information regarding simultaneous vicariance, with Tajima's D being the least correlated with the other summary statistics (Figs 4 and 5). By using statistical power to gauge the suitability of a particular summary statistic, our results indicate that $\text{var}(D)$, $\text{var}(\pi_{\text{net}})$, $\text{cov}(\pi_{\text{net}}, D)$, $\text{cov}(\pi, \pi_{\text{net}})$, $\text{cov}(\pi, D)$, and $\text{cov}(\pi_{\text{net}}, \theta_W)$ are useful measures for testing simultaneous vicariance. However, the power of these greatly depended on which hypothesis was being tested. When testing hypotheses of more recent divergence times (i.e. $H_{0.025A}$, $H_{0.25B}$, $H_{0.25A}$), $\text{var}(\pi_{\text{net}})$ achieved the highest statistical power whereas $\text{var}(D)$ was superior when testing older divergence time hypotheses (i.e. $H_{1.25A}$, $H_{1.25B}$, $H_{2.5A}$, $H_{2.5B}$, $H_{5.0A}$, H_{mixed} , H_{uniform} ; Table 3). This could be counter-intuitive because π_{net} is normally used to estimate divergence time (Nei & Li 1979) whereas significantly negative values of Tajima's D are often used to detect selective sweeps or population growth (Tajima 1989). However, the latter has a second purpose in being able to detect elevated balancing selection

or subdivision from significantly positive values. The superior power of Tajima's D at older near-simultaneous divergence times becomes clear with closer examination of our initial single taxon-pair simulations (Fig. 4) as well as previous theoretical studies (Takahata & Nei 1985; Hudson *et al.* 1987; Tajima 1989; Fu 1995; Wakeley & Hey 1997; Vasco *et al.* 2001; Pluzhnikov *et al.* 2002).

In our initial single taxon-pair simulations and consistent with previous studies, π_{net} increases with increasing τ (Fig. 4), and the greater accumulation of fixed differences between diverged populations leads Tajima's D to eventually increase with greater τ (Takahata & Nei 1985; Hudson *et al.* 1987; Simonsen *et al.* 1995; Fig. 4). Also consistent with previous studies (Takahata & Nei 1985; Hudson *et al.* 1987; Tajima 1989), the per-taxon-pair variance of π_{net} becomes greatly inflated with increasing τ (Fig. 4a), whereas the per-taxon-pair variance of Tajima's D is stable and extremely low in comparison (< 1.0 ; Fig. 4b). Therefore correctly rejecting older simultaneous vicariance histories should become difficult with $\text{var}(\pi_{\text{net}})$ because this test statistic becomes inflated across taxon pairs regardless of whether τ was nearly simultaneous ($H_{5.0A}$) or not (H_{mixed}). Even collecting more loci does not markedly erase this problem for $\text{var}(\pi_{\text{net}})$, in which the variance of π_{net} approaches 0.3 at the highest τ given 15 loci (Fig. 4h), whereas the variance of Tajima's D is approximately 0.05 throughout all τ given 15 loci (Fig. 4i).

Although we specifically consider the variance of a summary statistic (or covariance of pairs of summary statistics) across taxon pairs as a single test statistic for testing a comparative vicariance hypothesis, an ABC framework can use more information from the data by simultaneously using multiple summary statistics to test such hypotheses and estimate parameters (Estoup *et al.* 2004). Our simulation study suggests that such an ABC method should consider π , Tajima's D , π_{net} , and θ_W from each taxon pair as separate summary statistics, and especially Tajima's D because it was found to be relatively uncorrelated with the other summary statistics across divergence time (Fig. 5). In practice such a pool will likely be reduced because D and π_{net} are in fact composite summary statistics. In this case, we can consider their components as summary statistics, and again use these values from each taxon pair as separate summary statistics instead of the overall variance. For example, instead of using $\text{var}(\pi_{\text{net}})$, we could use each within taxon average pairwise difference (π_w) and average pairwise differences between each sister taxon (π_b). Likewise, instead of using Tajima's D , the summary statistic that was most uncorrelated with the other summary statistics, we can separately use its components, π , θ_W and $\text{var}(\pi - \theta_W)$ from each taxon pair as summary statistics (Tajima 1989). In this case, a data set of i taxon pairs would yield a pool of summary statistics that include $\pi_1 \dots \pi_i$, $\pi_{w1} \dots \pi_{wi}$, $\pi_{b1} \dots \pi_{bi}$, $\theta_{w1} \dots \theta_{wi}$, and $\text{var}(\pi - \theta_W)_1 \dots \text{var}(\pi - \theta_W)_i$.

Power and loci

Planning a phylogeographic study involves balancing the costs of an optimized sampling strategy with the resulting gains in statistical power for testing competing hypotheses. In such studies, increasing sample size is achieved by collecting data from more unlinked loci, base pairs, individuals and/or locations. In general, increasing sample size will reduce the variance in parameter estimates used to test hypotheses, yet these different strategies for increasing sample size have different statistical gains and drastically different monetary and temporal costs. For instance, collecting additional intron loci can be difficult given the obstacles of resolving phase and the difficulty in using 'universal primers' that are workable in a variety of taxa (Hare & Palumbi 2001; Zhang & Hewitt 2003; Aitken *et al.* 2004). It is then imperative to assess the gains in statistical power resulting from these various strategies for increasing 'sample size' before embarking on such a study. Although the resulting gains in power from collecting more base pairs or individuals are somewhat straightforward (Tajima 1983; Pluzhnikov & Donnelly 1996), determining how many unlinked nuclear loci are required for reasonable power in comparative phylogeographic studies is less obvious.

According to the properties of sample means, collecting additional loci should lower the variance in a summary statistic and this decrease in variance should be proportional to how many loci are collected. In the case of a summary statistic such as Tajima's D , collecting additional loci with equal mutation rates will result in $\text{var}(D) = \text{var}(D)/l$, where l is the number of loci (DeGroot 1986). This is largely manifest in our initial single taxon-pair simulations in which the variance of Tajima's D is 0.92 with a divergence time of zero ($\tau = 0$) and a single locus and 0.0612 with $\tau = 0$ and 15 loci collected (Fig. 4).

Under the multiple taxon-pair histories, uniformity in divergence times and other key parameters across taxon pairs will result in a further reduction in the variance of Tajima's D with more loci because the variance in D will be $\text{var}(D)/IT$, where T is number of taxa and l is the number of loci. Therefore the variable- τ history should consistently inflate $\text{var}(D)$ when collecting more loci given the effect that τ has on D averaged across loci demonstrated in our initial single taxon-pair simulations (Fig. 4). In this case $\text{var}(D)$ should be consistently inflated to the extent to which τ varies among taxa. It is encouraging that a single metric like $\text{var}(D)$ can successfully test simultaneous vicariance hypotheses with a moderate number of loci. Using the components of $\text{var}(D)$ simultaneously with the components of $\text{var}(\pi_{\text{net}})$ within a future ABC framework is likely to make testing simultaneous vicariance tractable given this reasonable number of loci and the range of parameter conditions we explored, instead of the nearly intractable problem presented by Edwards & Beerli (2000).

Edwards and Beerli proposed $l\tau/\theta > 1$ as a criterion for determining the number of loci required to reject $\tau = 0$ in a single taxon pair. Although it is a different hypothesis test, the power of $\text{var}(\pi_{\text{net}})$ to reject $H_{0.025A}$ given that $H_{0.25B}$ was the true history across 10 taxon pairs was > 0.9 given 5 loci, as opposed to the ~ 16.6 loci required to reject $\tau = 0$ at a single taxon pair according to their criterion (Table 3a). For many other scenarios, statistical power was > 0.5 with only 5–10 nuclear loci, an amount of data that can be feasibly collected in many laboratories, especially with the increasing phylogenetic breadth of genomic resources and development of efficient assays. While this does not conflict with the single taxon-pair criterion proposed by Edwards and Beerli, our success in being able to correctly reject simultaneous τ across 10 taxon pairs given among-taxon-pair variation in θ is in contrast to their pessimism regarding this difficult 'data-hungry' problem (Edwards & Beerli 2000).

Comparative phylogeographic complexity

It is sobering to note that some conditions will make testing simultaneous vicariance extremely difficult and will indeed make it a 'data-hungry' problem without having good prior information about how parameters vary across taxon pairs (Edwards & Beerli 2000). According to our simulations that incorporated post-divergence migration (results not shown), and previous theoretical work (Wakeley 1996; Nielsen & Slatkin 2000; Kalinowski 2002), migration after divergence will obscure the ability to distinguish among histories because it tends to erase the signal of genetic divergence between populations. Variance in demographic expansion could also hinder testing simultaneous vicariance. Although we found that a simple demographic expansion (Fig. 1) and a range of such demographic expansions after vicariance (results not shown) did not greatly hinder testing simultaneous vicariance, variance in demographic expansion across taxon pairs could exacerbate testing simultaneous vicariance by causing an elevation in the variance of Tajima's D under both simultaneous and variable divergence time histories (Simonsen *et al.* 1995). Furthermore, although we allowed θ to vary across taxon pairs, we did not explore variation in μ across loci, a factor that might greatly complicate such hypothesis testing. This reminds us that comparative phylogeographic studies will be more tractable in groups of taxa having low dispersal potential across the barrier leading to vicariance and when using loci for which the distribution of mutation rates across taxa or loci can be independently estimated.

Fortunately, an important advantage of simulation-based frameworks is that arbitrary complexity can be easily built into the underlying model. One can initially use simple models to identify processes that violate the simple model's assumptions. Subsequently, one can easily expand the simulation model to incorporate sufficient biological realism

and complexity that is guided by a priori hypotheses and independent estimates (Tavaré *et al.* 1997; Beaumont *et al.* 2002; Estoup *et al.* 2004). For example, independent estimates of variability in mutation rates can be directly incorporated by using a gamma-distributed prior distribution for μ . Alternatively, this flexibility allows one to incorporate uncertainty in across-loci variation in μ by using uninformative priors for the two free parameters of a gamma distribution (α and β). Likewise, uncertainty regarding variation in θ and demographic expansion parameters that vary across taxon pairs can be built into the model. Additionally, selection at certain loci can be detected and incorporated into a simulation-based analysis by hierarchical Bayesian methods or outlier tests (Lewontin & Krakauer 1973; Galtier *et al.* 2000; Storz & Beaumont 2002; Luikart *et al.* 2003; Beaumont 2004; Hey & Nielsen 2004). Although we did not report results incorporating recombination, we found that recombination only served to increase statistical power by decoupling linked polymorphisms and thereby reducing the variance (results not shown; Wakeley & Hey 1997).

Another consideration is that our model does not explicitly incorporate subdivision, a parameter that can strongly influence the patterns of the coalescent (Arbogast *et al.* 2002; Rosenberg & Feldman 2002). However, our coalescent model may still be justified if θ is rescaled by the number of demes and degree of subdivision, a reasonable assumption across various subdivision models (Wakeley 2000, 2004; Wakeley & Aliacar 2001; Nordborg & Krone 2002). Although the model we consider is not spatial, the general framework can be designed for geographically explicit models, such as comparative phylogeographic dispersal histories (Estoup *et al.* 2004) and ecologically deterministic distributional histories (Ravelo *et al.* 2004). In this case, the parameters underlying the biogeographic hypotheses that are based on ecological and palaeoclimatic models (Hugall *et al.* 2002) can be directly incorporated into the simulation model as parameters underlying the comparative phylogeographic vicariance hypotheses (Fig. 2).

Conclusion

We found that the power of a particular summary statistic used alone to test for simultaneous vicariance depended on which particular hypothesis was being tested. Using the summary statistics π , π_w , π_b , θ_w and $\text{var}(\pi - \theta_w)_1$ from each taxon pair simultaneously in an ABC framework should improve upon the statistical powers we report in this study, and therefore we are currently developing such a framework for comparative phylogeographic inference. Although we did not consider every summary statistic under the sun, we did investigate the commonly used ones that are known to capture the relevant signals underlying

comparative phylogeographic data sets with regards to demography and population divergence.

Acknowledgements

We thank S. Baird, M. Slatkin, R. Young, X. Zhao, F. Rousset and the anonymous reviewers for advice on improving this study. We thank J. Mackenzie for assisting us with the AWT data. We greatly thank E. Stahl, N. Takebayashi, K. Thornton, J. Novembre, and E. Anderson for assistance and advice regarding the simulations. Support for M. J. Hickerson was provided by a National Science Foundation postdoctoral grant in interdisciplinary informatics. The Perl and C code routines used for the simulations are available from M. Hickerson upon request.

References

- Aitken N, Smith S, Schwartz C, Morin PA (2004) Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology*, **13**, 1423–1431.
- Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.
- Avice JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.
- Avice JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 1707–1712.
- Bahlo M, Griffiths RC (2000) Inference from gene trees in a subdivided population. *Theoretical Population Biology*, **57**, 79–95.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–743.
- Beaumont BA (2004) Recent developments in genetic data analysis: what can they tell us about human demographic history? *Heredity*, **92**, 365–379.
- Beaumont BA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA*, **98**, 4563–4568.
- Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology*, **7**, 367–369.
- Brown JH (1995) *Macroecology*. University of Chicago Press, Chicago.
- Case TJ (1983) Sympatry and size similarity in *Cnemidophorus*. In: *Lizard Ecology: Studies of a Model Organism* (eds Huey RB, Pianka ER, Schoener TW), pp. 297–325. Harvard University Press, Cambridge, Massachusetts.
- Coates AG, Jackson JBC, Collins LS *et al.* (1992) Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin*, **104**, 814–828.
- Darwin C (1859) *The Origin of Species*. John Murray, London.
- DeGroot MH (1986) *Probability and Statistics*. Addison-Wesley, Reading, Massachusetts.

- Dolman G, Moritz C (in review) Demography of diversification of *carlia* skinks from Australian tropical rainforest: inference from multi-locus coalescent analyses.
- Dolman G, Phillips B (2004) Single-copy nuclear DNA markers characterized for comparative phylogeography in Australian wet tropics rainforest skinks. *Molecular Ecology Notes*, **4**, 185–187.
- Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.
- Estoup A, Wilson IJ, Sullivan CJ-MC, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, **159**, 1671–1687.
- Estoup A, Beaumont BA, Sennedot F, Moritz C, Cornuet J-M (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.
- Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506–509.
- Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169**, 1727–1738.
- Fu Y-X (1995) Statistical properties of segregating sites. *Theoretical Population Biology*, **48**, 172–197.
- Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
- Galtier N, Depaulis F, Barton NH (2000) Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics*, **2**, 981–987.
- Hare MP, Palumbi SR (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **16**, 700–706.
- 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, Cunningham CW (2005) Contrasting Quaternary histories in an ecologically divergent pair of low-dispersing intertidal fish (*Xiphister*) revealed by multi-locus DNA analysis. *Evolution*, **59**, 344–360.
- Hickerson MJ, Gilchrist MA, Takebayashi N (2003) Calibrating a molecular clock from phylogeographic data: moments and likelihood estimators. *Evolution*, **57**, 2216–2225.
- Hoelzer GA (1997) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution*, **51**, 622–626.
- Hudson RR (1983) Properties of a neutral model with intragenic recombination. *Theoretical Population Biology*, **23**, 183–201.
- Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds Futuyma D, Antonovics J), pp. 1–44. Oxford University Press, Oxford.
- Hudson RR (2002) ms – a program for generating samples under neutral models. *Bioinformatics*, **18**, 337–338.
- Hudson RR, Turelli M (2003) Stochasticity overrules the ‘three-times rule’: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.
- Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution. *Genetics*, **116**, 153–159.
- Hugall A, Moritz C, Moussalli A, Stanicic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences, USA*, **99**, 6112–6117.
- Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. *Molecular Biology and Evolution*, **15**, 1481–1490.
- Johnson NK, Cicero C (2004) new mitochondrial DNA data affirm the importance of Pleistocene speciation in North American Birds. *Evolution*, **58**, 1122–1130.
- Jordan DS (1908) The law of geminate species. *American Naturalist*, **42**, 73–80.
- Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. *Molecular Ecology*, **11**, 1263–1273.
- Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.
- Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 2257–2263.
- Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.
- Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution*, **55**, 955–975.
- Lewontin RC, Krakauer JK (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Luikart G, England PR, Tallman D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.
- MacArthur RH, Wilson EO (1967) *Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey.
- Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.
- Marko PB (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution*, **19**, 2005–2021.
- Moore WS (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, **49**, 718–726.
- Nei M, Li W (1979) Mathematical model for studying variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA*, **76**, 5269–5273.
- Nelson G, Platnick NI (1981) *Systematics and Biogeography: Cladistics and Vicariance*. Columbia University Press, New York.
- Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, **16**, 358–364.
- Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. *Evolution*, **54**, 44–50.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Nordborg M, Krone SM (2002) Separation of time scales and convergence to the coalescent in structured populations. In: *Modern Developments in Theoretical Population Genetics* (eds Slatkin M, Veuille M), pp. 130–164. Oxford University Press, Oxford.
- Plagnol V, Tavaré S (2002) Approximate Bayesian computation and MCMC. In: *Monte Carlo and Quasi-Monte Carlo Methods* (ed. Neiderreiter H), pp. 99–114. Springer-Verlag, Berlin.
- Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics*, **144**, 1247–1262.

- Pluzhnikov A, Di Rienzo A, Hudson RR (2002) Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics*, **161**, 1209–1218.
- Pritchard JK, Seielstad MTAP-L, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 1791–1798.
- Ravelo AC, Andreasen DH, Lyle M, Lyle AO, Wara MW (2004) Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature*, **429**, 263–267.
- Ricklefs RE, Schluter D (1993) Species diversity: regional and historical influences. In: *Species Diversity in Ecological Communities* (eds Ricklefs RE, Schluter D), pp. 350–363. University of Chicago Press, Chicago.
- Riddle BR, Hafner DJ, Alexander LF, Jaeger JR (2000) Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences, USA*, **97**, 14438–14443.
- Rosenberg NA, Feldman MW (2002) The relationship between coalescence times and population divergence times. In: *Modern Developments in Theoretical Population Genetics* (eds Slatkin M, Veuille M), pp. 130–164. Oxford University Press, Oxford.
- Schneider CJ, Cunningham M, Moritz C (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology*, **7**, 487–498.
- Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*, **141**, 413–429.
- Storz JF, Beaumont BA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Takahata N (1989) Gene genealogy in three related populations – consistency probability between gene and population trees. *Genetics*, **122**, 957–966.
- Takahata N, Nei M (1985) Gene genealogy and variance of intrapopulation nucleotide differences. *Genetics*, **110**, 325–344.
- Tallmon DA, Luikart G, Beaumont BA (2004) Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics*, **167**, 977–988.
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.
- Vasco DA, Crandall KA, Fu Y-X (2001) Molecular population genetics: coalescent methods based on summary statistics. In: *Computational and Evolutionary Analysis of HIV Molecular Sequences* (eds Rodrigo AG, Learn GH Jr). Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Wakeley J (1996) Distinguishing migration from isolation using the variance of pairwise differences. *Theoretical Population Biology*, **49**, 369–386.
- Wakeley J (2000) The effects of subdivision on the genetic divergence of populations and species. *Evolution*, **4**, 1092–1101.
- Wakeley J (2004) Recent trends in population genetics: More data! More math! Simple models? *Journal of Heredity*, **95**, 397–405.
- Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905.
- Wakeley J, Hey J (1997) Estimating ancestral population parameters. *Genetics*, **145**, 847–855.
- Watterson GA (1975) On the number of segregating sites in genetic models without recombination. *Theoretical Population Biology*, **7**, 256–276.
- Weiss G, von Haeseler A (1998) Inference of population history using a likelihood approach. *Genetics*, **149**, 1539–1546.
- Zhang D, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

M. Hickerson uses multi-locus DNA sequence data to study the biogeographic history of rocky intertidal fishes and additionally develops analytical methods for testing comparative phylogeographic hypotheses. As part of her PhD research, G. Dolman uses multi-locus DNA sequence data to study speciation and diversification processes in tropical rainforest fauna. C. Moritz is director of the museum of vertebrate zoology at the University of California, Berkeley, and his research program combines phylogeographic data with other historical and/or current distributional data to infer population processes in space and time.
