MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 4270–4282 doi: 10.1111/j.1365-294X.2009.04327.x

An information-theoretical approach to phylogeography

BRYAN C. CARSTENS, HOLLY N. STOUTE and NOAH M. REID

Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA

Abstract

Data analysis in phylogeographic investigations is typically conducted in either a qualitative manner, or alternatively via the testing of null hypotheses. The former, where inferences about population processes are derived from geographical patterns of genetic variation, may be subject to confirmation bias and prone to overinterpretation. Testing the predictions of null hypotheses is arguably less prone to bias than qualitative approaches, but only if the tested hypotheses are biologically meaningful. As it is difficult to know a priori if this is the case, there is the general need for additional methodological approaches in phylogeographic research. Here, we explore an alternative method for analysing phylogeographic data that utilizes information theory to quantify the probability of multiple hypotheses given the data. We accomplish this by augmenting the model-selection procedure implemented in IMA with calculations of Akaike Information Criterion scores and model probabilities. We generate a ranking of 17 models each representing a set of historical evolutionary processes that may have contributed to the evolution of *Plethodon idahoensis*, and then quantify the relative strength of support for each hypothesis given the data using metrics borrowed from information theory. Our results suggest that two models have high probability given the data. Each of these models includes population divergence and estimates of ancestral θ that differ from estimates of descendent θ, inferences consistent with prior work in this system. However, the models disagree in that one includes migration as a parameter and one does not, suggesting that there are two regions of parameter space that produce model likelihoods that are similar in magnitude given our data. Results of a simulation study suggest that when data are simulated with migration, most of the optimal models include migration as a parameter, and further that when all of the shared polymorphism results from incomplete lineage sorting, most of the optimal models do not. The results could also indicate a lack of precision, which may be a product of the amount of data that we have collected. In any case, the information-theoretic metrics that we have applied to the analysis of our data are statistically rigorous, as are hypothesis-testing approaches, but move beyond the 'reject/fail to reject' dichotomy of conventional hypothesis testing in a manner that provides considerably more flexibility to researchers.

Keywords: Coalescent, hypothesis-testing, Information theory, phylogeography

Received 23 March 2009; revision received 10 June 2009; accepted 28 June 2009

Geographical patterns of intraspecific variation have long been central to evolutionary biology (Wallace 1858; Darwin 1859; Mayr 1942), and descriptive figures that summarize such patterns of genetic variation are clearly valuable to researchers [e.g. Fig. 3 of Avise *et al.* (1987) and its imitators]. However, inferences regarding the evolutionary processes that have produced a particular pattern of genetic variation can be complicated by several factors. They may be subject to confirmation bias, a phenomenon where people interpret novel information in a manner that is consistent with their preconceived ideas (Nickerson 1998), and are also thought to be prone to over-interpretation (Knowles & Maddison 2002). Additionally, any inference about a given species is inherently dependent upon the capacity of individual researchers to envision realistic and meaningful historical demographic processes.

Correspondence: Bryan C. Carstens, Fax: 225 578 2597;
E-mail: carstens@lsu.edu

In response to the ubiquity of qualitative phylogeographic studies at the turn of the century, several researchers promoted the statistical testing of a priori hypotheses as an alternative to descriptive data analysis. For example, Sullivan *et al.* (2000) tested the predictions of hypotheses pertaining to the response of Mesoamerican rodents to climatic fluctuations using statistical approaches such as parametric bootstrapping rather than qualitative comparisons of tree topology. (Knowles (2001) made explicit predictions concerning the structure of Pleistocene refugia, and tested these predictions using simulations conducted under a coalescent model. Nielsen & Wakeley (2001) developed MDIV in large part to differentiate a model of population isolation strictly due to genetic drift from a model that included divergence with gene flow. The development of these methods, and others like them, was concurrent with the realization that estimates of many of the parameters that are important to phylogeography may contain a large amount of variance (Smouse 1998; Edwards & Beerli 2000; Arbogast *et al.* 2002; Hey & Machaho 2003; Hudson & Turelli 2003). In phylogeography, statistical approaches to hypothesis testing were designed to minimize inference error by accounting for the statistical error inherent to phylogenetic parameter estimation (Hickerson & Cunningham 2005), as well as the stochasticity associated with the process of allelic coalescence (Knowles & Maddison 2002). Accounting for both types of error is necessary for statistical tests of phylogeographic hypotheses (Carstens *et al.* 2005a). However, while a rigorous statistical framework for testing a priori phylogeographic hypotheses has theoretical advantages, it can be difficult to apply in systems that lack the extrinsic information required to develop hypotheses. Hypotheses are most easily developed in systems with external information pertaining to the historical range of a species, for example those with fossil or paleoenvironmental data (Cruzan & Templeton 2000; Brunhoff *et al.* 2003; Cognato *et al.* 2003; Tribsch & Schonswetter 2003; McCormack *et al.* 2008), or in specific habitats that shift in predictable ways with climate change (Knowles 2001; DeChaine & Martin 2005). While paleoenvironmental niche modelling can aid in the development of hypotheses in systems impacted by recent environmental events (Carstens & Richards 2007; Richards *et al.* 2007), many researchers seek to investigate systems where, for a variety of reasons, a priori hypotheses are not readily developed. Furthermore, the intrinsic utility of null hypothesis testing is dependent upon the quality of the hypotheses; rejecting a hypothesis that does not capture meaningful information concerning the biology of the system provides nothing of value to a researcher (Knowles 2004). For these reasons, the community of phylogeographers would benefit from considering alternatives to descriptive and hypothesis testing approaches.

Phylogeographic research seeks to identify historical processes that have been important to the evolution of the focal taxon and to identify how these processes have contributed to the formation of population genetic structure. As such, phylogeography is clearly a historical rather than an experimental science. While descriptive investigations are widely used across the historical sciences and are valuable, approaches that seek to identify the hypothesis that best explains the historical data are desirable (Cleland 2001). For example, Chamberlin (1890) advocated a method of testing multiple working hypotheses in the historical sciences where several plausible hypotheses are evaluated on the basis of the strength of the evidence in their favour. Phylogeography is a discipline well-suited to Chamberlin's approach, because any combination of a number of historical processes (e.g. population subdivision, population divergence, population size change, recombination, migration/hybridization) may have contributed to the evolution of a given species. As the primary objective of any phylogeographic investigation is to identify the relative contributions of these processes, we could conceptualize our discipline as one that seeks to identify the most appropriate model of population demographic history for the focal taxon because the selection of this model will aid us in identifying the historical processes that have been most important in its evolution. Adopting this approach requires a subtle reversal in the relationship between hypotheses and data, in comparison to that used in testing null hypotheses.

When null hypotheses are tested in phylogeography, the *P*-value represents the probability that the test statistic (which is estimated from the empirical data) is greater than expected given that the null hypothesis is true (e.g. Sokal & Rohlf 1994). However, if the data are collected to evaluate multiple working hypotheses, it is desirable to reverse this relationship and calculate the probability of each hypothesis given the data (Anderson 2008). Reversing the relationship between data and hypotheses has an additional benefit to the historical sciences because it allows a set of n models (representing n hypotheses) to be ranked. In this manner, the probability of the hypothesis given the data can be calculated for any number of hypotheses, these probabilities can be compared and evaluated using approaches from information theory (Kullback 1959). In contrast, if we use *P*-values to test the set of n hypotheses, we would reject some portion and fail to reject the others, but would be unable to differentiate among those that could not be rejected.

An information-theoretical approach to phylogeographic research would proceed by ranking a set of hypotheses in terms of their probability given the data, rather than describing patterns in the data and deriving hypotheses from these patterns or calculating the probability of some parameter while assuming a particular hypothesis is true. In information-theoretical approaches, *information* is anything that decreases our uncertainty about the relative contributions of historical processes (Burnham & Anderson 1998; Anderson 2008). In the context of phylogeographic research, historical processes (represented by parameters) can be incorporated into a mathematical model and can be compared by calculating the likelihood of a model given the data. Thus, any software for genetic data analysis that calculates the likelihood of the full model and also allows users to calculate this likelihood using only a subset of the total parameters in the model could potentially be used for an information-theoretical approach. In fact, Nielsen & Wakeley (2001) used an information-theoretic statistic to evaluate models when they introduced MDIV; basically they calculated Akaike Information Criterion (AIC) scores for models that included and did not include gene flow (Nielsen & Wakeley 2001). Akaike (1973) defined $\text{AIC} = -2\log(\text{model}\,|\,\text{data}) + 2k$, where k is the number of estimated parameters in the model. By taking the exponential of the difference between the AIC score of the best model and the AIC score of model i, and the weighting across all models, the model probability (w_i) of model i given the data can be calculated and compared to other models (Anderson 2008). Any number of models can be ranked using AIC scores, and further this approach is not limited to comparisons of a single nested model to a full model, as are likelihood ratio tests.

Until recently, one major impediment to the application of information-theoretical approaches has been the requirement that users conduct a complete run for each possible submodel—for example, one MDIV analysis without migration, and one with migration for the question explored by Nielsen & Wakeley (2001). However, a recently introduced extension of the isolation-with-migration model, IMA (Hey & Nielsen 2007), includes a novel procedure for model selection. The model implemented in IMA assumes two descendant and one ancestral population, and estimates θ from each population, migration rates between each descendant population, as well as the time since population divergence. IMA utilizes an computational approach first described by (Kuhner *et al.* 1995) to estimate the joint posterior probabilities of the model parameters, and it is these estimates that enable researchers to evaluate nested demographic models. As the likelihoods of each of 16 reduced models (Fig. 1) given the data are calculated, users of IMA are also able to assess the statistical significance of biologically important parameters such as gene flow, divergence among populations, and population size by applying information-theoretical approaches to model selection.

To our knowledge, two investigations have used some type of model-selection approach with the isolation-with-migration model (Brumfield *et al.* 2008; Geraldes *et al.* 2008). Brumfield *et al.* (2008) used IMA to compare

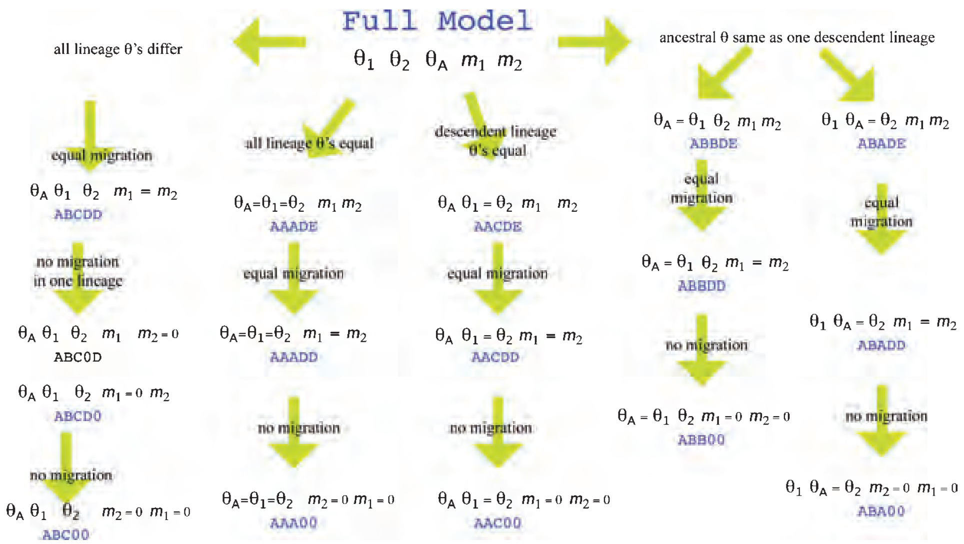

Fig. 1 Schematic of the relationships among submodels in IMA. Shown is the full model, which estimates five parameters, which are subsequently reduced in various combinations in each of the submodels. Note that IMA also estimates population divergence (t), which is not included in the model-selection procedure because t is not included in the joint posterior density function of IMA (J. Hey, personal communication: http://groups.google.com/group/Isolation-with-Migration).

a model where migration was symmetrical and population θs were equal ($\theta_1 = \theta_2 = \theta_A$ and $m_{12} = m_{21}$) for several pairwise comparisons of Amazonian birds. Geraldes *et al.* (2008) compared models with and without gene flow to test hypotheses related to introgression of the Y-chromosome in European rabbits. Both studies were motivated by the desire to ascertain the biological importance of the various parameters; for example, was it justifiable to treat population θs as equal (Brumfield *et al.* 2008), or has gene flow historically occurred at a high level (Geraldes *et al.* 2008). Given that many investigations use the isolation-with-migration model to explore gene flow between divergent populations (Nikula *et al.* 2007; Kotlik *et al.* 2008; Lee & Edwards 2008; McCormack *et al.* 2008; Niemiller *et al.* 2008; Szovenyi *et al.* 2008), as well as introgression/hybridization (Carling & Brumfield 2008; Good *et al.* 2008; Hird & Sullivan 2009), the model-selection approach implemented within IMA has broad application to phylogeography. However, the greatest value of this approach is that it allows phylogeographic researchers to evaluate multiple working models in the spirit of Chamberlin (1890), thereby quantifying the relative strength of support for various combinations of historical processes. Rather than describing the pattern of genetic variation or testing a priori hypotheses, researchers can identify processes that have been important in structuring the genetic diversity within a given species by evaluating competing models using approaches such as AIC (Burnham & Anderson 1998). Here, we collect data from a representative empirical system, analyse these data and evaluate models using IMA, explore the performance of the model selection using simulated data, and provide an example of how information theory can be used to identify hypotheses that are well-supported given the data.

Methods

Empirical data

Over 250 *Plethodon idahoensis* were collected from northern Idaho, Montana, and British Columbia as part of a broader comparative phylogeographic investigation into the Pacific Northwest mesic forest ecosystem (Carstens *et al.* 2004, 2005b). Data from the cytochrome *b* (cyt *b*) mitochondrial gene suggest that significant population structure within the species is found between the two southern-most river drainages (the Lochsa and Selway Rivers) and the remaining northern drainages. Descriptive evidence supports this finding, for example there is monophyly of individuals from the southern drainages, and results from an analysis of molecular variance (Carstens *et al.* 2004) suggest that southern populations have historically been isolated from northern populations. Coalescent-based tests of null hypotheses have also been conducted (Carstens *et al.* 2005b; Carstens & Richards 2007), and generally these tests support the idea that populations were restricted to multiple glacial refugia during the Pleistocene. Furthermore, there is evidence for restricted gene flow between these populations (Carstens *et al.* 2005c) and population expansion (Carstens *et al.* 2004). Consequently, the *P. idahoensis* system provides an opportunity to explore the utility of an information-theoretical approach to phylogeography using the model-selection procedure implemented in IMA because the set of parameters included in this model encompass the two most relevant questions to this system: (i) has migration between populations historically occurred at high levels between the southern and northern groups within *P. idahoensis* and (ii) has there been a substantial change in the size of the ancestral populations (as would be predicted under a refugial model). In *P. idahoensis*, answering these questions will improve our understanding of how historical demographic processes that occurred as a result of Pleistocene climatic fluctuations.

To explore these questions, we add data from two nuclear genes to a previously collected 669-bp fragment of cyt *b* (Carstens *et al.* 2004). Additional data were collected from the internally transcribed spacer of the 16SrRNA gene (Hillis & Dixon 1991) and from the recombination activating protein (RAG) 1 gene (Wiens *et al.* 2006). We gathered data from 30 *P. idahoensis*, twelve individuals from the Lochsa or Selway drainages, and eighteen from the North Fork of the Clearwater and other northern drainages. Samples were chosen at random from available tissue samples, not on the basis of information from the previously collected cyt *b* data. In developing this sampling strategy, we sought to balance the anticipated improvement in accuracy of the parameter estimates that usually accompanies an increase in the number of loci (Edwards & Beerli 2000; Carling & Brumfield 2007; Carstens & Knowles 2007) with the computational demands of large multilocus data sets. In choosing three loci and thirty individuals, we hope to collect enough data to estimate parameters reasonably well in an amount of time that is short enough to allow a sensitivity analysis to be conducted using simulated data and replication (details of this analysis below). Sequence data were phased to alleles either by polymerase chain reaction (PCR) subcloning (using high fidelity *Taq*) and subsequently sequencing eight clones from each sample, or through the use of PHASE (Stephens *et al.* 2001).

The isolation-with-migration model in IMA assumes that loci are evolving in a neutral manner, and also that there is no recombination within each locus. To explore these assumptions, we calculated Tajima's (1989) *D* in each of the populations. While a significantly negative value of *D* can result from natural selection or from demographic effects (such as population expansion or structure), values that are not significant suggest to us that these loci are not inappropriate to use in an IMA analysis. However, this test may have low power given our samples sizes (36 and 24 chromosomes as collected here in the northern and Lochsa/Selway populations) and subsequently the protein-coding genes used in this analysis could be subject to weak purifying selection that may not be detected by Tajima's *D* (Hammer *et al.* 2003). We also conducted the four-gamete test in each population using DnaSP (Rozas & Rozas 1999) to test for recombination.

Gene flow, population divergence, and $\theta = 4N_e\mu$ were estimated using IMA (Hey & Nielsen 2007) for all the data. Initially, 25 runs of IMA were conducted with varying prior values for the theta, migration and divergence time parameters with a fixed run length of 25 h and six coupled Markov chains. Once optimal priors were identified, we experimented by varying the number of coupled Markov chains (2–10) and the heating scheme. Via this process, we settled on prior values of $\theta_1 = 25$, $\theta_2 = 25$, $\theta_A = 50$, $m_{12} = 1$, $m_{21} = 1$, $t = 10$, and determined that four coupled Markov chains and a geometric heating with parameters that varied between $-g_1 = 0.05$ and $-g_2 = 10.0$. We then conducted two sets of IMA runs with varying run times (25, 50 and 100 h) on a 3.0 GHz MacPro, using random number seeds for each and a burn-in period of 100 000 steps. In these runs, we analysed our data using the full isolation-with-migration model and conducted the model-selection procedure independently on each run to explore the effects of run-time variation on the selection procedure. The posterior density function of the full isolation-with-migration model is maximized to generate parameter estimates in IMA (Hey & Nielsen 2007) and the model-selection procedure repeats this process for each of 16 reduced models (Fig. 1). Due to reasons discussed by Hey & Nielsen (2007), the population divergence parameter (*t*) is not included in this procedure. Hey & Nielsen (2007) conduct model selection by comparing the results of the likelihood ratio test (LRT) to an appropriate chi-squared distribution, and also imply that a correction for multiple comparisons is appropriate. As a result, we conducted model selection using both uncorrected LRTs and a Bonferroni correction for multiple comparisons. We also computed the AIC scores (Akaike 1973), AIC differences, model likelihoods and probabilities following Anderson (2008).

Simulation study

Once we were confident that we had good estimates of parameters from our empirical data, we used MS (Hudson 2002) to simulate genealogies for each locus under three coalescent models. The first of these corresponded to the full isolation-with-migration model (ABCDE), with values of θ_1, θ_2, θ_A, m_{12}, m_{21}, and *t* that matched those estimated from our empirical data. The second model did not include migration (ABC00), but was simulated using values of θ_1, θ_2, θ_A, and *t* that matched those from our empirical data. The third model also did not include migration (AAA00), and did not incorporate differences in θ among the ancestral and descendant populations (e.g. $\theta_1 = \theta_2 = \theta_A$ and *t*). For the third model, $\theta_1 = \theta_2 = \theta_A$ was set to the arithmetic mean of the empirical estimates. For each set of genealogies, we used SEQ-GEN (Rambaut & Grassly 1997) to simulate sequence data with characteristics (model of sequence evolution, number of segregating sites) matching our empirical data. Once these data were simulated, we assembled input files for IMA using UNIX and PERL scripts. In this way, we were able to produce data that either did not differ appreciably from the empirical data (the full model), or differed in known ways (the *no migration* or *no migration with equal theta* models). These simulated data were then analysed using the model-selection protocol implemented in IMA in order to explore the sensitivity of the program. We conducted 100 replicates, including the model-selection protocol, for each model.

Results

Analysis of empirical data

Sequence data were collected from 30 individuals, including 884 bp from the ITS of the 16s rRNA gene, and 1210 bp from the RAG-1 gene and are deposited in GenBank under Accession numbers GQ337923–GQ337954 (ITS) and GQ247792–GQ247811 (RAG-1). Sequence data were resolved to alleles using either PCR subcloning ($n_{\text{ITS}} = 13$; $n_{\text{Rag1}} = 12$) or PHASE. While levels of polymorphism in these genes were not as high as in the mitochondrial DNA, they nevertheless exhibited considerable variation (Table 1). Results from the Tajima's *D*-test (Table 1) and the four-gamete test suggest that the assumptions of the isolation-with-migration model are not violated by these data. Sequence alignment of all loci was unambiguous and conducted manually.

Our preliminary analysis suggested that optimal mixing occurred with four coupled Markov chains and a geometric heating scheme ($-g_1 = 0.05$ and $-g_2 = 2.0$). All metrics suggested that the Markov chains converged before the end of the burn-in and were adequately sampling the posterior distribution of parameter space. For example, the lowest ESS value, which was recorded

Table 1 Sequence data statistics

Gene	bp	*s*	*h*	π	θ_w	Fixed	*D*	*P*-value
Cyt *b*								
All	669	27	13	0.00822	0.01093	2	−0.9358	>0.10
LS		10	6	0.0045	0.0041	0	0.3353	>0.10
ND		15	13	0.00474	0.0060	0	−0.7415	>0.10
ITS								
All	884	14	25	0.00482	0.00390	0	0.7483	>0.10
LS		13	15	0.00536	0.00422	0	0.9210	>0.10
ND		9	16	0.00337	0.00273	0	0.7727	>0.10
Rag1								
All	1210	13	17	0.00217	0.00253	0	−0.4347	>0.10
LS		8	8	0.0016	0.00188	0	−0.4918	>0.10
ND		8	11	0.00773	0.00183	0	−0.1751	>0.10

Characteristics of sequence data 675 for three loci are shown, for 676 all the data as well as the Lochsa/Selway (LS) and northern drainages (ND) populations. 677 Characteristics include the length (bp), number of segregating sites (*s*), heterozygosity (*h*), 678 nucleotide diversity (π), Waterson's theta (θ_w), the number of fixed differences between 679 populations, Tajima's *D* and its *P*-value.

during one of the 25-h runs, was ESS = 99, and for longer runs this value was considerably higher (e.g. ESS = 563 for one of the 100-h runs). Similarly, there was no discernable pattern to the plots of parameter trend lines. After examining the IMA output of all runs, we are convinced that there is no evidence that suggests that the Markov chain have not converged by the end of the burn-in period.

Estimates of parameters from the IMA analysis (Table 2) have at least three biologically important implications. First, estimates of ancestral θ are lower than estimates of descendent θ, this could indicate isolation in a small ancestral population. Second, estimates of divergence time between populations are much greater than zero, which supports earlier findings of population structure. Mean migration rates, while small, are nonzero, implying that divergence has occurred with gene flow. One plausible interpretation of values of this magnitude is population divergence with low rates of gene flow. Since these parameters are reported in terms of coalescent units, conversion to values that are more easily interpreted requires some assumptions about the mutation rate, which we can not calculate directly. However, divergence between *Plethodon idahoensis* and its sister taxon *Plethodon vandykei* is probably between 5 and 2 Myr (Carstens *et al.* 2005b; Wiens *et al.* 2006); given this value the temporal divergence between the Lochsa/Selway populations and the northern drainages is between 27 000 and 68 000 generations, and migration rates are something on the order of one individual per 1000 generations. This suggests that there is not a strong pattern of ongoing gene flow between these populations.

Table 2 Parameter estimates from the empirical data using IMA

	θ_1	θ_2	θ_A	m_1	m_2	t
HiPt						
Run1	9.5137	8.8014	1.5263	0.0005	0.0295	1.905
Run2	9.4119	8.8014	1.5263	0.0005	0.0245	1.905
Mean						
Run1	10.892	9.7872	5.0007	0.1382	0.1166	2.2543
Run2	10.8813	9.7741	5.3667	0.1406	0.1156	2.2874
$\text{HPD}_{90\text{Lo}}$						
Run1	5.6472	5.3419	0.1018	0.0005	0.0005	1.195
Run2	5.6472	5.3419	0.1018	0.0005	0.0005	1.175
$\text{HPD}_{90\text{Hi}}$						
Run1	15.924	13.9907	5.5963	0.3055	0.2455	3.025
Run2	15.924	13.9907	5.7998	0.3125	0.2445	3.075

For each parameter, the high point, mean, and boundaries of the highest posterior density are given. Results of two long runs are shown.

Results of model selection across each of the runs were similar, and for ease of discussion we combined the posterior distributions from the two long runs (corresponded to more than 3.5×10^8 steps in the Markov chain) and repeated the model-selection procedure (Table 3). These results illustrate the complex nature of the joint estimates of model parameters. For example, using an uncorrected likelihood ratio test, there are nine models that can be rejected and seven which can not be rejected. Other than models that force population θ values to be equal (e.g. AAA– models; which were all rejected), it is difficult to draw generalities about which parameters are 'important' to include in a model of *P. idahoensis* population demography from these results. For example, some of the models which can not be rejected do not include a migration parameter, while others which can not be rejected include one or two migration parameters. Further, some of the models

Table 3 Results of the model selection using IMA

Model	t	$-\log(P)$	θ_1	θ_2	θ_A	m_1	m_2	d.f.	$-\ln L$	2LLR	w/out BC	w/BC
AAC00	1.6713	2.6181	10.8797	10.8797	2.4262	0.0001	0.0001	3	4382.5391	1.7462	No	No
AACDD	2.1643	1.8577	9.7182	9.7182	1.5016	0.0577	0.0577	2	4381.7787	0.2254	No	No
ABC00	1.6749	2.5452	11.7299	10.2152	2.4266	0.0001	0.0001	2	4382.4662	1.6004	No	No
ABC0D	2.1027	1.7472	10.3274	9.4342	1.5316	0.0001	0.095	1	4381.6682	0.0044	No	No
AACDE	2.103	1.7736	9.8597	9.8597	1.5307	0.0035	0.0872	1	4381.6946	0.0572	No	No
ABCDD	2.1697	1.8449	9.8631	9.4589	1.4751	0.058	0.058	1	4381.76585	0.1997	No	No
ABCD0	2.0076	2.2723	10.3021	9.7436	1.8714	0.1131	0.0001	1	4382.1933	1.0546	No	No
AAA00	1.283	6.9995	9.1936	9.1936	9.1936	0.0001	0.0001	4	4386.92045	10.5089	**Yes**	No
ABB00	1.351	6.0887	11.9582	7.3859	7.3859	0.0001	0.0001	3	4386.00965	8.6873	**Yes**	No
AAADD	1.777	6.1114	8.096	8.096	8.096	0.0713	0.0713	3	4386.03235	8.7327	**Yes**	No
ABBDE	1.6291	5.3206	11.1877	6.8544	6.8544	0.000101	0.1088	1	4385.2416	7.1512	**Yes**	No
ABBDD	1.7295	5.454	10.5582	6.93	6.93	0.0635	0.0635	2	4385.375	7.418	**Yes**	No
ABA00	1.33	6.7374	8.2059	10.4829	8.2059	0.0001	0.0001	3	4386.65835	9.9847	**Yes**	No
ABADD	1.7694	5.8735	7.2364	9.3331	7.2364	0.068	0.068	2	4385.7945	8.257	**Yes**	No
AAADE	1.7766	6.1114	8.0709	8.0709	8.0709	0.0704	0.072	2	4386.03235	8.7327	**Yes**	No
ABADE	1.7186	5.8333	7.0689	9.9757	7.0689	0.1598	0.000126	1	4385.7542	8.1764	**Yes**	No

Shown are values from the first long IMA run of the empirical data. High point values for each parameter are shown, in addition to the degrees of freedom for each model, marginal likelihoods of each model given the data, likelihood ratio test statistic (2LLR), and the results with and without a Bonferroni correction (BC) for multiple comparisons. Note that estimates of $m_1 = 0.0001$ and $m_2 = 0.0001$ are indistinguishable from zero.

which can not be rejected allow θ to differ among all lineages, while others allow only the descendant and ancestral θ to differ. Most (but not all) of the models which can not be rejected share two general patterns: they either have low values of ancestral θ_A (~2.3), higher values for descendent θ s (~11), and no migration, or are those in which the respective θ values are even lower (θ_A~1.5 and $\theta_{1,2}$~9.5) with some migration ($m_{1,2}$~0.05). In short, if a model with migration is fitted to the data, the estimates of ancestral and descendant θ are reduced. Accompanying these differences is a difference in the estimate of population divergence (t); models with migration tend to have larger estimates of t than models without (Table 3). Note that if we correct for multiple comparisons using a Bonferroni correction (e.g. by dividing 0.05 by 16), we can not reject any of the models.

While we can learn something useful about evolutionary processes that have been important in the *P. idahoensis* system from the likelihood ratio tests, these tests do not differentiate among the models which can not be rejected, and as such do not allow identification of the optimal model. An information-theoretical approach to phylogeography requires some objective way of ranking these models, so we computed AIC scores (Akaike 1973) for each model (Table 4). The AIC is based on both the probability of the model given the data and the number of parameters such that, should two models have the same likelihood and differ in the number of parameters, the model with fewer parameters will have a better AIC. When models were ranked by AIC, the best model was one that included $\theta_1 = \theta_2$, θ_A, and t (AAC00). The second-best model included migration $\theta_1 = \theta_2$, θ_A, $m_{12} = m_{21}$ and t (AACDD). We quantified the difference between models by comparing the model with the best AIC score to any other model ($\Delta i = \text{AIC}_i - \text{AIC}_{\min}$). The resulting Δi values estimate

Table 4 Information-theoretic statistics for each of the IMA models

Model	k	AIC	Δ_i	w_i	$E_{\min,i}$
AAC00	2	8769.0782	0	0.471458846	1
AACDD	3	8769.5574	0.4792	0.291964382	1.61478206
ABC00	3	8770.9324	1.8542	0.073820156	6.386586937
ABC0D	4	8771.3364	2.2582	0.049285593	9.565855122
AACDE	4	8771.3892	2.311	0.046750821	10.08450412
ABCDD	4	8771.5317	2.4535	0.04054173	11.62897699
ABCD0	4	8772.3866	3.3084	0.017243441	27.34134431
FULL	5	8773.332	4.2538	0.006699493	70.37231972
ABB00	1	8775.8409	6.7627	0.000545055	864.9744769
ABA00	2	8776.0193	6.9411	0.000455997	1033.906887
ABBDD	2	8776.0647	6.9865	0.000435758	1081.928093
AAA00	3	8776.4832	7.405	0.000286743	1644.184836
ABADD	3	8776.75	7.6718	0.000219595	2146.942466
AAADD	2	8777.3167	8.2385	0.000124597	3783.860259
ABBDE	3	8777.589	8.5108	9.49E-05	4968.136008
ABADE	3	8778.0647	8.9865	5.90E-05	7994.427376
AAADE	4	8779.5084	10.4302	1.39E-05	33867.12301

Shown are models considered by IMA, the number of parameters for each model, its AIC score, AIC differences (Δi), model probabilities (w_j) and evidence ratio ($E_{\min,j}$). All values were calculated following Burnham & Anderson (1998).

Kullback & Leibler (1951) information (e.g. the distance between the best model and model *i*), and are easy to interpret; models improve relative to the best model as Δi approaches zero. The AIC Δi between the two best models (AAC00 and AACDD) is $\sim$0.5, and other models are substantially worse. Two other statistics are used to transform AIC values into metrics that are more easily interpreted. We calculated Akaike weights (w_i), the normalized relative likelihoods of the models given the data, and evidence ratios ($w_{\min}/w_i$) following Burnham & Anderson (2002). Akaike weights provide a way to quantify how much of the total evidence favours a particular model—for example, some 76% of the total likelihood is contributed by the AAC00 and AACDD models. Evidence ratios provide a way to conceptualize the relative odds that a model is best given the data. If the Akaike weights of model *i* and the best model are 0.12 and 0.6, respectively, then the evidence ratio is 0.6/0.12 = 5, and the odds against model *i* being the best model are 5:1. When applied to our data, evidence ratios suggest that the second-best model (AACDD) has 1.6:1 odds compared to the best model (AAC00), a degree of support that we consider sufficient to consider strongly (Table 4).

Table 5 Results from the simulation study

IMA model	Uncorrected			Corrected multiple comparisons		
	ABCDE	ABC00	AAA00	ABCDE	ABC00	AAA00
AAADE	0.03	0.04	0.05	0.01	0.00	0.01
AAADD	0.10	0.06	0.08	0.01	0.00	0.01
AAA00	0.54	0.00	0.08	0.12	0.00	0.00
ABC00	0.61	0.06	0.11	0.19	0.00	0.00
ABCD0	0.30	0.10	0.04	0.09	0.00	0.00
ABC0D	0.40	0.03	0.07	0.12	0.00	0.00
ABCDD	0.03	0.00	0.00	0.00	0.00	0.00
ABADE	0.11	0.09	0.05	0.01	0.00	0.00
ABADD	0.06	0.00	0.01	0.00	0.00	0.00
ABA00	0.59	0.01	0.05	0.15	0.00	0.00
ABBDE	0.06	0.03	0.03	0.01	0.00	0.00
ABB00	0.56	0.01	0.05	0.18	0.00	0.00
AACDE	0.09	0.03	0.05	0.01	0.00	0.01
AACDD	0.03	0.03	0.03	0.00	0.00	0.01
AAC00	0.59	0.01	0.09	0.15	0.00	0.00
AAADE	0.03	0.04	0.05	0.01	0.00	0.01

For each model, the proportion of replicates where the model could be rejected using a likelihood ratio test, for each of three sets of simulated data corresponding to the ABCDE, ABC00, and AAA00 models are shown. Columns on the left are not corrected for multiple comparisons, columns on the right are corrected using a Bonferroni correction.

Simulation study

Simulated data were analysed in order to evaluate the sensitivity of the model-selection procedure in IMA given parameter values similar to those observed in our empirical data. With one exception (total run time), we used the same settings as in our empirical IMA runs. As results of the long runs of the empirical data were not appreciably different from the shorter runs, we set run length on the simulated data to 24 h, which corresponded to an average of 4.3×10^6 steps in the Markov chain for each replicate. Simulated data were analysed using the Louisiana State University High Performance Computing network on a cluster computer with multiple 2.33 GHz nodes.

Two important results follow from this sensitivity analysis. Results suggest that the procedure is sensitive to migration. When migration is present in the simulated data at levels commensurate with estimates from our empirical data, models that do not include migration as a parameter will typically be rejected (Table 5). However, when data are simulated without migration, models that do not include migration as a parameter are not always rejected, regardless if θ is allowed to vary (Table 5), unless we correct for multiple comparisons. A correction for multiple comparisons is appropriate here, as we are making 16 comparisons of the likelihood of model *i* to the likelihood of the full model, and have applied a Bonferroni correction to these results (Table 5). This correction clarified our results to some extent; the suggestion that models that do not include migration as a parameter are easily rejected when migration has been a historically important process is supported. However, when data are simulated with a constant θ, these differences are not sufficient to reject models where θ is allowed to vary. We interpret these results to mean that the model-selection procedure in IMA is more sensitive to violations involving the migration parameters than the θ parameters for values similar to those found in our empirical data.

Results from the simulated data analysis are more meaningful when AIC values are calculated and an information-theoretical approach is used to interpret the results. We calculated AIC values for all models across replicates, and determined that in nearly all cases the model with the best AIC score captured some meaningful information about the conditions that the data were simulated under, even if the best model did not match the simulation model exactly. For example, when data were simulated with migration and θs that varied by lineage, a model with at least one migration parameter and two θ parameters was chosen in 70% of the replicates (Table 6). When data were simulated either with no migration, or with no migration and equal θs, models that captured a similar proportion of the meaningful

Table 6 Models selected for the simulated data replicates

θ_1, θ_2, θ_A, m_{12}, m_{21}(ABCDE)		θ_1, θ_2, θ_A, $m_{12} = 0$, $m_{21} = 0$(ABC00)		$\theta_1 = \theta_2 = \theta_A$, $m_{12} = 0$, $m_{21} = 0$(AAA00)	
ABCD0	21	AAC00	36	AAA00	40
AACDD	17	ABB00	21	AAC00	20
ABC0D	15	ABA00	20	ABA00	14
AAC00	14	ABC0D	8	ABB00	13
ABB00	9	ABCD0	4	ABADD	3
ABCDE	7	AACDD	3	ABC0D	3
ABBDD	6	ABADD	2	ABCD0	2
ABADD	5	ABBDD	2	AACDD	2
ABAD0	3	AACDE	1	ABADE	1
AACDE	2	ABADE	1	ABBDD	1
ABBDE	1	ABBDE	1	ABC00	1
ABC00	1	ABC00	1		

A count of the number of times that various models had the best AIC scores for the simulation study. All models ranked by the number of times that had the best AIC score, for each of the three sets of simulated data corresponding to the ABCDE, ABC00 and AAA00 models are shown.

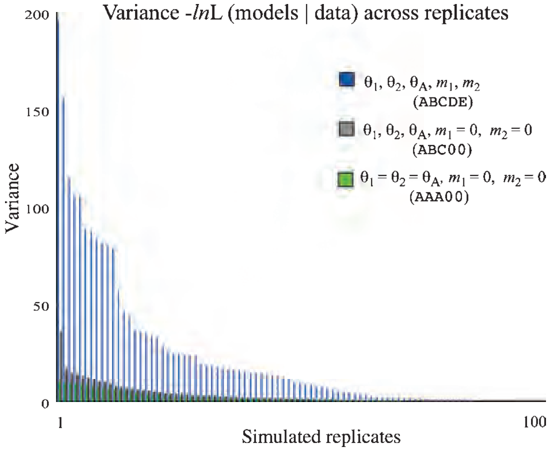

Fig. 2 Variance in the likelihood of the models given the data. Each bar represents the variance in the marginal likelihoods across the sixteen models considered by IMA, for data simulated under one of three conditions. Blue bars correspond to data simulated with migration and with θs that varied by lineage, grey bars correspond to data simulated without migration but with θs that varied by lineage, and green bars correspond to data simulated without migration and with θs that were equal among lineages. Note that some bars on the right-tail of each distribution are too small to visualize on the scales of the *y*-axis.

parameters were chosen in 89% and 81% of the replicates, respectively. These results are probably influenced by the number of processes that have shaped the evolutionary dynamics of a given system. For example, it may be more costly to under parameterize models when data are simulated with migration and θs that differ across lineages, than it is to over parameterize models when data are simulated without migration and with θs that are the same across lineages. This suggestion is supported by the dramatic difference in the variance in the marginal likelihoods of the models given the data across the three sets of simulated data (Fig. 2).

Discussion

Empirical investigation

Information-theoretic metrics allow us to make rigorous and unbiased inferences about the recent historical demography of *Plethodon idahoensis*. We move beyond the 'reject/fail to reject' approach inherent to null hypothesis testing, by following Chamberlin (1890) in treating each of the models within IMA as a hypothesis that describes a particular set of historical processes that may have influenced the evolution of *P. idahoensis*. Calculating AIC scores allows us to quantify the strength of support in the data for various models (Anderson 2008). Curiously, the two models that are most probable given the data are biologically very different. One of these (AAC00) includes migration as a parameter, and the other (AACDD) does not (Table 4). This suggests that we have two regions of parameter space with a high probability given the data, one that includes migration and one that does not. Each of these regions includes estimates of ancestral and descendent θ, as well as population divergence (*t*), but the estimates of these parameters vary (Table 3). When migration is included in the model, estimates of *t* increase and estimates θ of decrease relative to the model that does not include migration. While in each case the estimates of θ_A are small relative to the estimates of the descendant θ (supporting the hypothesis that *P. idahoensis* was isolated in multiple Pleistocene refugia), the absolute values of these parameters change depending on the inclusion of the migration parameter.

Results from the empirical data suggest that the parameter space describe by the IMA model is complex, and that observed patterns in the data can be explained by more than one combination of parameters. Apart from evaluating the relative plausibility of a set of hypotheses, information-theoretic evidence ratios provide an efficient way to explore the effects of adding data. For example, one could explore the effects of sequentially adding loci or samples to the evidence ratios, which could either be accomplished via power analyses with simulated data or by subsampling from the empirical data. AIC scores support the interpretation that we have two competing models of historical demography that are highly probable given the data (AAC00 and AACDD). To some degree, these models each are supported by specific patterns in each locus; the models with the highest AIC scores for the ITS, Rag1 and cyt *b* loci were the ABCDE, AAC00, and AAC00, respectively. When data from ITS/cyt *b* and ITS/Rag1 were combined, the AACDD model had the best AIC value, and when data from Rag1/cyt *b* were combined, the AAC00 model was best.

Simulation study

As the model-selection approach included in IMA has to date only been used on a handful of systems, we used simulated data to assess the performance of the model-selection procedure given our data. Our results (Table 5) suggest that the procedure implemented in IMA is able to reject models without migration when migration has been historically important in a large proportion of the replicates. Data that were simulated without migration, but with different population sizes, were more problematic for model selection (Table 5). We suspect that some combination of the following explanations can account for these results. First, data were simulated under values of t, θ, and m matching our empirical data, and it is possible, particularly in the case of the θ_1 and θ_2 values, that these were too similar to one another to easily differentiate the AAC– from the ABC– models. Second, our run times for the simulation study were shorter than those used in the empirical study. While ESS values suggest that the Markov chains had reached stationarity, it may be that our sample of the posterior distribution is inadequate to compute the marginal likelihoods of the models accurately. To explore this issue, we repeated five of the ABCDE runs while doubling the run time, and obtained similar results, which suggested to us that our run times were adequate for this simulation study. Finally, it is conceivable that we have not collected data from a sufficient number of loci, or from a large enough number of individuals, to actually differentiate the AAC00 from the AACDD and other models.

Are information-theoretical approaches broadly useful to phylogeography?

Phylogeographic research aims to understand the unique combination of historical processes that have shaped the evolution of the focal taxon (Avise *et al.* 1987). The isolation-with-migration model is increasingly important to the discipline, as evidence by the over 300 citations of the manuscripts that introduced IM and IMA since 2004 (Web of Science). This suggests to us that the information-theoretical approach described above should be broadly useful, even if it can not currently be applied to every phylogeographic system because these systems do not match the model used in IMA. Examples of inappropriate systems include those where one of the descendent lineages is founded by a small subset of the ancestral population (a splitting parameter is included in IM but not IMA) and those that are divided into more than two populations. It is also worth noting that this approach can be extended to questions beyond those that fit the isolation-with-migration model. For example, the computer program LAMARC allows users to include or exclude parameters such as migration and population size change in its calculation of the relative likelihood of the model given the data (Kuhner 2006), and as such could be applicable to phylogeographic systems with more than two populations to evaluate the relative contributions of gene flow and population expansion if these relative likelihoods were comparable. Other questions are amenable to this information-theoretic framework; for example Fitzpatrick *et al.* (2008) used Bayes factors to compare two models implemented in Structure to see if they could reject a model of hybridization at a contact zone (Pritchard *et al.* 2000; Fitzpatrick *et al.* 2008). In essence, approaches that seek to identify the best demographic model given the data are similar to the model-selection procedures used by phylogenetists to identify models of sequence evolution (e.g. Sullivan & Joyce 2005), and in general model-selection approaches are broadly applicable across ecology and evolutionary biology (Johnson & Omland 2004). While no current method implements that phylogeographic equivalent of the GTR+I+Γ model of sequence evolution, such a model is theoretically possible. Alternatively, methods such as approximate Bayesian computation could provide phylogeographers with the flexibility to evaluate a wide range of hypotheses (Nielsen & Beaumont 2009). While certain biological processes (e.g. gene flow, demographic expansion, genetic drift) are more easily incorporated into a likelihood framework than others (e.g. selection), we nevertheless remain hopeful that upcoming methological advances will allow us to consider more complex models. Certainly the release of IMA 2.0 (J. Hey, personal communication) while this manuscript was in revision is a positive sign that more complex models can be evaluated with information-theoretical approaches.

Phylogeographic research seeks to identify the historical processes that have influenced genetic variation within populations. In practice, many investigators gathered data in order to develop or test hypotheses that are essentially models of historical population demography. For example, Carstens *et al.* (2005b) posited that isolation in multiple refugia during Pleistocene glacial periods, followed by postglacial expansion, was a likely explanation for the pattern of genetic diversity observed in *P. idahoensis*. This scenario is essentially a model of the historical demography, albeit one that is rather short on specifics, that has been developed over the course of several publications. While it may be unfair to characterize this approach as laborious, it necessarily involves a number of different types of analyses, each with some specific aim. Will phylogeography proceed in this manner in the future, as sequence data become increasingly easier to collect, as multilocus data sets grow into genomic data sets, and as more complex analytical methods become available? We hope not. Rather, we envision a future where a structured coalescent model can incorporate *all* of the historical processes that may influence genetic diversity, where one can estimate parameters associated with these models, and one where phylgeographic researchers apply information-theoretical approaches to evaluate the strength of support for particular hypotheses contained within the full model. This approach would be as broadly applicable as the conventional descriptive approach, but at least as rigorous as statistical approaches to hypothesis testing.

Acknowledgements

This research was supported in part by a Howard Hughes Medical Institute grant through the Undergraduate Biological Sciences Education Program to Louisiana State University to H.N. Stoute and by the Louisiana Board of Regents who provided N.M. Reid with a Doctoral Fellowship. Many of the analyses were conducted using computational resources provided by the Louisiana Optical Network and by the Louisiana State University High Performance Computing system. We thank J.D. McVay, M.M. Koopman, and T. Pelletier for providing comments on drafts of this manuscript, and J.K. Carstens for assistance collecting the sequence data. We are grateful to participants in the 2009 AMNH-SWRS Statistical Phylogeography workshop for their extensive discussions related to this topic. We thank Rémy Petit and four anonymous referees for their helpful comments and suggestions on the initial submission of this manuscript.

References

Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory* (eds Petrov BN, Csaki F), pp. 267–281. Akademiai Kiado, Budapest.

Anderson DR (2008) *Model Based Inference in the Life Sciences*. Springer, New York.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski J (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, **33**, 707–740.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews of Ecology, Evolution, and Systematics*, **18**, 489–522.

Brumfield RT, Liu L, Lum DE, Edwards SV (2008) Comparison of species tree methods for reconstructing the phylogeny of Bearded Manakins (Aves: Pipridae, *Manacus*) from multilocus sequences data. *Systematic Biology*, **57**, 719–731.

Brunhoff C, Galbreath E, Federov VB, Cook JA, Jaarola M (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957–968.

Burnham KP, Anderson DR (1998) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel inference: A practical Information-Theoretic Approach*. Springer, New York.

Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity. *PLoS ONE*, **e160**, 1–6.

Carling MD, Brumfield RT (2008) Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. *Genetics*, **178**, 363–377.

Carstens BC, Knowles LL (2007) Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, **16**, 619–627.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J (2004a) Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Systematic Biology*, **53**, 781–792.

Carstens BC, Degenhardt JD, Stevenson AL, Sullivan J (2005a) Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. *Molecular Ecology*, **14**, 255–265.

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J (2005b) Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**, 1639–1652.

Carstens BC, Bankhead A III, Joyce P, Sullivan J (2005c) Testing population genetic structure using parametric bootstrapping. *Genetica*, **124**, 71–75.

Chamberlin TC (1890) The method of multiple working hypotheses. *Journal of Geology*, **5**, 937.

Cleland CA (2001) Historical science, experimental science, and the scientific method. *Geology*, **29**, 987–990.

Cognato AI, Harlin AD, Fisher ML (2003) Genetic structure among Pinyon Pine Beetle populations (Scolytinae: *Ips confusus*). *Environmental Entomology*, **32**, 1262–1270.

Cruzan MB, Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution*, **15**, 491–496.

Darwin C (1859) *On the Origin of Species. A Facsimile of the First Edition*. Harvard University Press, Cambridge, MA.

DeChaine EG, Martin AP (2005) Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. *Journal of Biogeography*, **32**, 1943–1956.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescent time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Fitzpatrick BM, Placyk JS Jr, Niemiller ML, Casper GS, Burghardt GM (2008) Distinctiveness in the face of gene flow: hybridization between specialist and generalist gartersnakes. *Molecular Ecology*, **17**, 4107–4117.

Geraldes A, Carneiro M, Delibes-Mateos M *et al.* (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian peninsula. *Molecular Ecology*, **17**, 4489–4499.

Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, **17**, 1313–1327.

Hammer MF, Blackmer F, Garrigan D, Nachman MW, Wilder JA (2003) Human population structure and its effects on sampling Y chromosome sequence variation. *Genetics*, **164**, 1495–1509.

Hey J, Machaho CA (2003) The study of structured populations—new hope for a difficult and divided science. *Nature Reviews Genetics*, **4**, 535–543.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation or improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

Hickerson MJ, Cunningham CW (2005) Contrasting Quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus dna analysis. *Evolution*, **59**, 344–360.

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, **66**, 411–453.

Hird S, Sullivan J (2009) Assessment of gene flow across a hybrid zone in red-tailed chipmunks (*Tamias ruficaudis*). *Molecular Ecology*, **18**, 3097–3109

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

Hudson RR, Turelli M (2003) Stochasticity overrules the ''three-times rule'': genetic drift, genetic draft, and coalescent times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.

Johnson JB, Omland KE (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**, 101–108.

Knowles LL (2001) Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers. *Molecular Ecology*, **10**, 691–701.

Knowles LL (2004) The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, **17**, 1–10.

Knowles LL, Maddison DR (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Kotlik P, Markova S, Choleva L *et al.* (2008) Divergence with gene flow between Ponto-Caspian refugia in an anadromous cyprinid *Rutilus frissi* revealed by multiple gene phylogeography. *Molecular Ecology*, **17**, 1076–1088.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hasting sampling. *Genetics*, **140**, 1421–1430.

Kullback S (1959) *Information Theory and Statistics*. John Wiley, New York.

Kullback S, Leibler RA (1951) On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

Lee JY, Edwards SV (2008) Divergence across Australia's Carpentarian Barrier: statistical phylogeography of the Red-backed Fairy Wren (*Malurus melanocephalus*). *Evolution*, **62**, 3117–3134.

Mayr E (1942) *Systematics and the Origin of the Species*. Columbia University Press, New York, NY.

McCormack JE, Bowen BS, Smith TB (2008) Integrating paleoecology and genetics of bird populations in two sky island archipelagos. *BMC Biology*, **6**, 1–12.

Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, **2**, 175–220.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.

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

Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. *Molecular Ecology*, **17**, 2258–2275.

Nikula R, Stelkov P, Valnola R (2007) Diversity and trans-Arctic invasion history of mitochondrial lienages in the North Atlantic *Macoma balthica* complex (Bivalvia: Tellinidae). *Evolution*, **61**, 928–941.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rambaut A, Grassly NC (1997) SEQ-GEN: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computational Applications to Biosciences*, **13**, 235–238.

Richards CL, Carstens BC, Knowles LL (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographic hypotheses. *Journal of Biogeography*, **34**, 1833–1845.

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Smouse P (1998) To tree or not to tree. *Molecular Ecology*, **7**, 399–412.

Sokal RR, Rohlf FJ (1994) *Biometry: The Principles and Practices of Statistics in Biological Research*, 3rd edn. W. H. Freeman, New York.

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

Sullivan J, Joyce P (2005) Model selection in phylogenetics. *Annual Reviews of Ecology, Evolution, and Systematics*, **36**, 445–466.

Sullivan J, Arellano E, Rogers DS (2000) Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent response to past climatic fluctuations. *The American Naturalist*, **155**, 755–768.

Szovenyi P, Terracciano S, Ricca M, Giordano S, Shaw AJ (2008) Recent divergence, intercontinental dispersal and shared polymorphism are shaping the genetic structure of amphi-Atlantic peatmoss populations. *Molecular Ecology*, **17**, 5364–5377.

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **105**, 437–460.

Tribsch A, Schonswetter P (2003) Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon*, **52**, 477–497.

Wallace AR (1858) On the tendency of varieties to depart indefinitely from the original type. *Journal of the Proceedings of the Linnean Society: Zoology*, **3**, 53–62.

Wiens JJ, Engstrom TN, Chippindale PT (2006) Rapid diversification, incomplete isolation, and the ''speciation clock'' in North American salamanders (Genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution*, **60**, 2585–2603.

Bryan Carstens is interested in methodological approaches to phylogenetics and phylogeography, particularly in how the approaches that are adopted while analyzing phylogeographic data can influence our confidence in the results. Noah Reid is interested in phylogentic theory and comparative phylogeography of neotropical insects. Holly Stoute conducted this research while supported as a summer researcher by the HHMI; she is interested in evolutionary genetics and developmental biology, and hopes to enter a Ph.D. program in the Fall of 2010.

M E C 4 4 9 3 B
Journal Name Manuscript No.
Dispatch: 17.12.09 Journal: MEC CE: Shanthalakshmi
Author Received: No. of pages: 1 PE: Shyamala

Molecular Ecology (2009) doi: 10.1111/j.1365-294X.2009.04493.x

Corrigendum

Contributed by BRYAN C. CARSTENS, HOLLY N. STOUTE and NOAH REID
Department of Biological Sciences, 202 Life Sciences Building, Louisiana State University, Baton Rouge, LA 70803, USA

Carstens *et al*. (2009) presented a method for evaluating models of historical demography under the isolation-with-migration model implemented in the software package IMa (Hey & Nielsen 2007). Our manuscript suggested that phylogeographic inferences could be made with a demographic model-selection framework using AIC scores of the models and information theoretic statistics (Burnham & Anderson 1998). Regretfully, we made a mistake in the calculation of the $-\ln L$ values reported in Table 3 (Carstens *et al*. 2009). This mistake resulted from a misinterpretation of the IMa output; it is not possible to calculate the likelihood of the reduced models using IMa. However, we are still able to compare the relative AIC differences among models using the $-\log(P)$ values, which are the maximized posterior density functions of each of the models considered by IMa given the data. Their usage in the calculation of information theory metrics such as the AIC differences (Δ_i) follows from the justification provided by Hey & Nielsen (2007), who used these values to conduct likelihood ratio tests. While the absolute AIC values (reported in Table 4) are affected by the erroneous calculation of the likelihoods of the reduced models, the relative differences among AIC values do not change when the $-\log(P)$ values are used to calculate AIC scores. Consequently, neither the biological inferences regarding the evolution of *Pseudofagus idahoensis*, or the general approach to phylogeography suggested by Carstens *et al*. (2009) are compromised.

Table 4 Information theoretic statistics for each of the IMa models. Shown are models considered by IMa, the number of parameters for each model, its AIC score, AIC differences (Δ_i), model probabilities ($\mathbf{w}_i$) and evidence ratio ($E_{min,j}$). All values were calculated following Burnham & Anderson (1998)

Model	k	AIC	Δ_i	w_i	$E_{min,i}$
AAC00	2	9.2362	0	0.471458846	1
AACDD	3	9.7154	0.4792	0.291964382	1.61478206
ABC00	3	11.0904	1.8542	0.073820156	6.386586937
ABC0D	4	11.4944	2.2582	0.049285593	9.565855122
AACDE	4	11.5472	2.311	0.046750821	10.08450412
ABCDD	4	11.6898	2.4535	0.04054173	11.62897699
ABCD0	4	12.5446	3.3084	0.017243441	27.34134431
FULL	5	13.49	4.2538	0.006699493	70.37231972
ABB00	1	15.999	6.7627	0.000545055	864.9744769
ABA00	2	16.1774	6.9411	0.000455997	1033.906887
ABBDD	2	16.2228	6.9865	0.000435758	1081.928093
AAA00	3	16.6412	7.405	0.000286743	1644.184836
ABADD	3	16.908	7.6718	0.000219595	2146.942466
AAADD	2	17.4748	8.2385	0.000124597	3783.860259
ABBDE	3	17.747	8.5108	9.49E-05	4968.136008
ABADE	3	18.2228	8.9865	5.90E-05	7994.427376
AAADE	4	19.6666	10.4302	1.39E-05	33867.12301

References

Burnham KP, Anderson DR (1998) *Model Selection and Inference: A Practical Information–Theoretic Approach*. Springer, New York.

Carstens BC, Reid N, Stoute HN (2009) An information theoretical approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation or improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Correspondence: Bryan C. Carstens, Fax:???? ???? ????; E-mail: carstens@lsu.edu