City University of New York (CUNY)
CUNY Academic Works

Publications and Research | City College of New York

September 2014

Detecting Concerted Demographic Response across Community Assemblages Using Hierarchical Approximate Bayesian Computation

Yvonne L. Chan
University of Hawaii at Manoa

David Schanzenbach
University of Hawaii at Manoa

Michael J. Hickerson
CUNY City College

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/cc_pubs

Recommended Citation
Chan, Y. L., Schanzenbach, D. & Hickerson, M. J. (2014). Detecting Concerted Demographic Response across Community Assemblages Using Hierarchical Approximate Bayesian Computation. Molecular Biology and Evolution, 31(9), 2501-2515. doi:10.1093/molbev/msu187.

This Article is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.

Detecting Concerted Demographic Response across Community Assemblages Using Hierarchical Approximate Bayesian Computation

Yvonne L. Chan,*[†,1] David Schanzenbach,[2] and Michael J. Hickerson[3,4]

[1]Hawai'i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai'i at Manoa

[2]Cyberinfrastructure, University of Hawai'i at Manoa

[3]Biology Department, City College of New York

[4]The Graduate Center, City University of New York

†Present address: Department of Genetics and Bioinformatics, Swedish Museum of Natural History, Stockholm, Sweden

***Corresponding author:** E-mail: ylhchan@hawaii.edu.

Associate editor: Noah Rosenberg

Abstract

Methods that integrate population-level sampling from multiple taxa into a single community-level analysis are an essential addition to the comparative phylogeographic toolkit. Detecting how species within communities have demographically tracked each other in space and time is important for understanding the effects of future climate and landscape changes and the resulting acceleration of extinctions, biological invasions, and potential surges in adaptive evolution. Here, we present a statistical framework for such an analysis based on hierarchical approximate Bayesian computation (hABC) with the goal of detecting concerted demographic histories across an ecological assemblage. Our method combines population genetic data sets from multiple taxa into a single analysis to estimate: 1) the proportion of a community sample that demographically expanded in a temporally clustered pulse and 2) when the pulse occurred. To validate the accuracy and utility of this new approach, we use simulation cross-validation experiments and subsequently analyze an empirical data set of 32 avian populations from Australia that are hypothesized to have expanded from smaller refugia populations in the late Pleistocene. The method can accommodate data set heterogeneity such as variability in effective population size, mutation rates, and sample sizes across species and exploits the statistical strength from the simultaneous analysis of multiple species. This hABC framework used in a multitaxa demographic context can increase our understanding of the impact of historical climate change by determining what proportion of the community responded in concert or independently and can be used with a wide variety of comparative phylogeographic data sets as biota-wide DNA barcoding data sets accumulate.

Key words: **comparative phylogeography, approximate Bayesian computation, historical demography, response to climate change.**

Article

Introduction

Fluctuations in climate during the Quaternary resulted in widespread expansions and contractions of ice sheets, regional shifts in temperature and precipitation, and changes in sea levels and sea surface temperatures (Yu and Eicher 1998; Cutler et al. 2003; Pahnke et al. 2003). These fluctuations brought about profound changes in population sizes, species range distributions, and geographic patterns of genetic diversity in nearly all species (Graham et al. 1996; Comes and Kadereit 1998; Dynesius and Jansson 2000; Hewitt 2000, 2004; Davis and Shaw 2001; Hill et al. 2011). Characterizing the complex dynamics and aggregate demographic histories associated with expansions and contractions of large species assemblages is a daunting yet vital component of learning how climatic change can drive patterns of meta-community assembly in various regions (HilleRisLambers et al. 2012).

Despite the large number of phylogeographic studies examining species response to historical changes that have emerged since the field's inception in the 1980s, there remains a pressing need for the integration of multispecies data sets within a cohesive statistical framework (Knowles 2009). Methods that use genetic data from whole assemblages can then be used to determine the community-level patterns of both individualistic and aggregate responses to climate cycles. Although community-level inference (Taberlet et al. 2002; Soltis et al. 2006; Drew and Barber 2012; Wood et al. 2012) from aggregate parameter estimates from each species and/or qualitative comparisons of phylogenetic trees or phylogeographic networks from each species can provide useful inferences about community responses to changes in climate and landscape (Sullivan et al. 2000; Soltis et al. 2006), a unified statistical framework that can pool information across species would be broadly beneficial.

In this study, we present a new and generally applicable multispecies population genetic method that is suitable for aggregate demographic histories of any community provided

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

the availability of population-level sampling of single locus data across codistributed species. This general method can potentially make use of the plethora of data from nearly 25 years of phylogeographic studies (Hickerson et al. 2010) and mitochondrial DNA (mtDNA) barcode initiatives (17,505 chordate species; Ratnasingham and Hebert 2007, as of May 2014).

Under this strategy, this multitaxon statistical framework makes community-level inference while accommodating coalescent, mutational, and demographic variance associated with individual species and loci (Beaumont 2010; Morgan et al. 2011) and gains statistical strength from pooling data for a single analysis (Beaumont and Rannala 2004). Specifically, our multitaxon statistical framework uses hierarchical approximate Bayesian computation (hABC). Under complex models that make calculating the likelihood function difficult, ABC uses a comparison of observed to simulated summary statistics to sidestep the calculation of the likelihood function and develop approximate likelihoods for model choice and parameters (Sunnåker et al. 2013). hABC is an ABC extension of any hierarchical Bayesian model where multiple parameters are structured into multiple dependent levels whereby observable outcomes are modeled conditional on certain parameters, which themselves are given probabilistic specification in terms of higher level parameters, known as hyperparameters (Gelman et al. 2004). Hierarchical models are particularly useful when there are a number of quantities such as loci or populations, and it is unknown whether they should be parameterized the same way or independently (Beaumont 2010).

However, up to now, most hABC studies in a comparative phylogeographic context focused on studies of codivergence (Hickerson and Meyer 2008; Carnaval et al. 2009; Barber and Klicka 2010; Morgan et al. 2011) and local adaptation (Bazin et al. 2010). The method we introduce here is the first multispecies coalescent model-based method for the demographic inference of coordinated and/or independent multipopulation expansion histories, thereby allowing large-scale inferences relevant to questions about community assembly and variable responses to future climate changes.

The hABC Approach

Approximate Bayesian computation is particularly well suited to address the complexity of communities and natural populations where evaluation of the likelihood function is difficult (Csilléry et al. 2010). Often based on the rejection algorithm (Tavare et al. 1997; Pritchard et al. 1999), it involves the simulation of large numbers of data sets under different hypothesized scenarios with parameter values drawn from a prior. Simulated and observed data sets are reduced to summary statistics and then posterior probability distributions are approximated from the comparisons between the simulated and observed summary statistics. Although it has been used frequently in population and evolutionary genetics to estimate species-specific parameters such as effective population size and past demographic events such as growth, decline, and migration under complex demographic histories (Beaumont 2010), its application to community-wide dynamics is relatively recent (Hickerson et al. 2006) where it has been used to test for simultaneous divergence among sister taxa. Here, we infer community-wide dynamics based on a demographic history of concerted response. Our goal was to provide a tool to determine what proportion of a community responded simultaneously to a hypothesized historical event, such as a sea level change or expansion out of refugia, and to estimate the timing of the response.

We present an hABC scheme for detecting the presence and timing of multispecies coexpansion pulses at the time scale of the late Quaternary (table 1). We harness this approach to infer the proportion of a contemporary community that coexpanded simultaneously, as well as when that expansion occurred. Rather than compiling and qualitatively comparing a set of single-species inferences, this hABC method allows estimation of hyperparameters that quantify multispecies patterns, such as synchronicity in expansion time. By combining the data in a single hierarchical Bayesian analysis, we incorporate uncertainty in the amount of dependency among taxon-specific parameters, thereby allowing for both historical congruence and demographic independence across taxa (Beaumont 2010).

First, we apply the hABC framework in extensive simulation experiments to validate the method and to examine the robustness of our model estimates and parameter estimation to different sizes of community samples and maximum expansion times. We introduce a model index hyperparameter of community congruence ζ, which ranges from complete simultaneous expansion ($\zeta = 1.0$), where all n species expanded at the same time τ_s to complete random expansion ($\zeta = 0.0$), where each of the n species expanded independently at times τ_i. For intermediate models, ζ is proportional to the number of species that coexpand synchronously with $(1-\zeta)$ being proportional to the number of species that expand individualistically at random times τ_i. All populations have a demographic model that consists of a contemporary effective size of N_i that is independently drawn from a uniform prior $U(N_{min}, N_{max})$, with every population instantaneously expanding from populations that are a fraction of their current size ε_i chosen from a uniform prior of expansion magnitudes $U(\varepsilon_{min}, \varepsilon_{max})$ at times τ_i or τ_s that have a uniform prior distribution of $U(\tau_{min}, \tau_{max})$ (fig. 1). Although each of the n taxa's effective population size N_i and expansion magnitude ε_i are free to vary, the method allows estimating the following: 1) ζ, the proportion of the n taxa that synchronously expand; 2) $E(\tau)$, the mean expansion time of n taxa; 3) τ_s, the coexpansion time of the synchronous taxa; and 4) the dispersion index of all n expansion times, $\mathrm{Var}(\tau)/E(\tau)$.

To demonstrate the method's bias and accuracy given that the model is correct, we performed a series of cross-validation tests by simulating data sets having fixed values for the community congruence hyperparameter index ζ (pseudo-observed data sets [PODs]; Bertorelle et al. 2010) and subsequently estimating ζ, $E(\tau)$, τ_s, and $\mathrm{Var}(\tau)/E(\tau)$ using the ABC accept–reject method. To investigate the method's performance when the model is incorrect (i.e., model generating the data differs from the model used for inference) and if such incorrect models could possibly be detected prior to

Table 1. Hierarchical Approximate Bayesian Computation Procedure Outlined for Comparative Phylogeographic Inference of Concerted Demographic Response.

Objectives	• Estimate the proportion of n taxa that coexpanded synchronously (ζ) • Estimate mean time of expansion $E(\tau)$, coexpansion time τ_s, and time-dispersion $\mathrm{Var}(\tau)/E(\tau)$
Draw hyperparameter value from prior	1) For each data set of n taxa, draw a value of ζ representing $\|\|n\zeta\|\|$ taxa that are coexpanding synchronously across n taxa
Draw taxon-specific parameter values from prior	2) Draw an expansion time τ_a and for each $\|\|1-n\zeta\|\|$ taxa expanding individualistically and a single τ_s from the prior for all $\|\|n\zeta\|\|$ taxa that are coexpanding synchronously. Assign each of the n taxa an effective population size (N), expansion magnitude (ε), and mutation rate (u) drawn from taxon-specific priors
Simulate data for each taxon and calculate summary statistics	3) Simulate data based on sample sizes, sequence lengths, and taxon-specific parameter values for all the n taxa using a coalescent simulation program 4) Record parameter values and calculate summary statistics (number of haplotypes, haplotypic diversity, nucleotide diversity, and Tajima's D) for each of the n taxa
Calculate hypersummary statistics	5) Calculate summary statistics for each set of n taxa based on the mean, variance, skewness, and kurtosis to compress multitaxon data set into D_j 6) Repeat steps 1–5 for X iterations 7) Compute hypersummary statistics from the observed multitaxon data set D^*
Accept/Reject	8) Accept D_j for 1,000 smallest Euclidean distances between simulated and observed$\|D_j-D^*\|$ and record hyperparameter values ζ, $E(\tau)$, τ_s, and $\mathrm{Var}(\tau)/E(\tau)$. Reject the remaining
Estimation	9) Fit local linear regression model to 1,000 accepted data sets and adjust hyperparameter values to obtain joint posterior probability estimates of $E(\tau)$, $\mathrm{Var}(\tau)/E(\tau)$,$\tau_s$, and ζ_k

applying our method, we simulated PODs from communities that differed from our coexpansion and random expansion community by including nonexpanding populations, declining populations, and multiple congruent pulses, and performed graphical checks using principal component analysis (PCA) given the PODs and data sampled from the hyperprior.

Finally, we apply this method to an empirical data set of 32 avian populations from Australia to determine how many bird species coexpanded simultaneously and the timing of any coexpansion. We validate our estimates with PODs using the sampling and species-specific prior distributions of these 32 avian populations and conduct graphical model checking of the prior and posterior predictive distributions. Thus, we verify that the model can generate the main features of the observed data. We then use PCA to help identify which of the 32 species coexpanded, followed by an hABC analysis on this subset to confirm species identification and validate the temporal coexpansion, and finally, to estimate when the coexpansion occurred.

Results

Cross-Validation of hABC Method

Sensitivity analyses to the size of the community (10 vs. 50 species) and the time frame (maximum expansion time $\tau_{\max} = 100{,}000$ or 500,000 years before present) show that, as expected, the estimates of mean time $E(\tau)$ and dispersion index of time ($\mathrm{Var}(\tau)/E(\tau)$) are substantially more accurate when a greater proportion of the taxa are coexpanding (fig. 2 and table 2). In addition, although the number of taxa sampled from a community (10 vs. 50) does have a strong effect on being able to test for synchronous or asynchronous coexpansion, the timeframe is much less so.

Model assumptions are an important component of all model-based methods (Csilléry et al. 2010), therefore we examined the behavior of three sets of PODs with key differences from the model used to generate prior samples contained in the reference table. These three sets of PODs were generated under the following three models: 1) "constant"—15 taxa expansion at the same time and 35 taxa with zero population growth; 2) "declining"—15 taxa expansion at the same time and 35 taxa declining to one-tenth their original size at random times; and 3) "two pulse"—15 taxa expansion at 30,000 generations before present and 35 taxa expansion at 90,000 generations before present. Using PCA, we compared data generated from each of these models to data generated from a model with 15 species coexpanding at a single time and 35 species expanding at independent times (i.e., $\zeta = 0.3$) and expansion magnitudes ε_i drawn from the uniform prior $U(\varepsilon_{\min}, \varepsilon_{\max})$.

Using PCA as a graphical check to compare the PODs and data generated from the prior, we found that data generated from the constant and declining models can be detected with PCA to be a poor fit for hABC analysis under our coexpansion model (fig. 3*A* and *B*). In contrast, this PCA technique suggested that data generated from the two-pulse model were not found to be distinguishable from data generated from our single pulse $\zeta = 0.3$ model (fig. 3*C*).

Pleistocene Expansion Times in Australian Avian Populations

To demonstrate the method on an empirical data set, we applied this hABC method to data from 32 Australian bird populations. For this analysis, we specifically chose species in which all samples represented a monophyletic cluster lacking in genetic structure and putatively expanded from a single refugium. After using PCA to conduct graphical model checks of the prior and posterior predictive distributions to verify that the model can generate the main features of

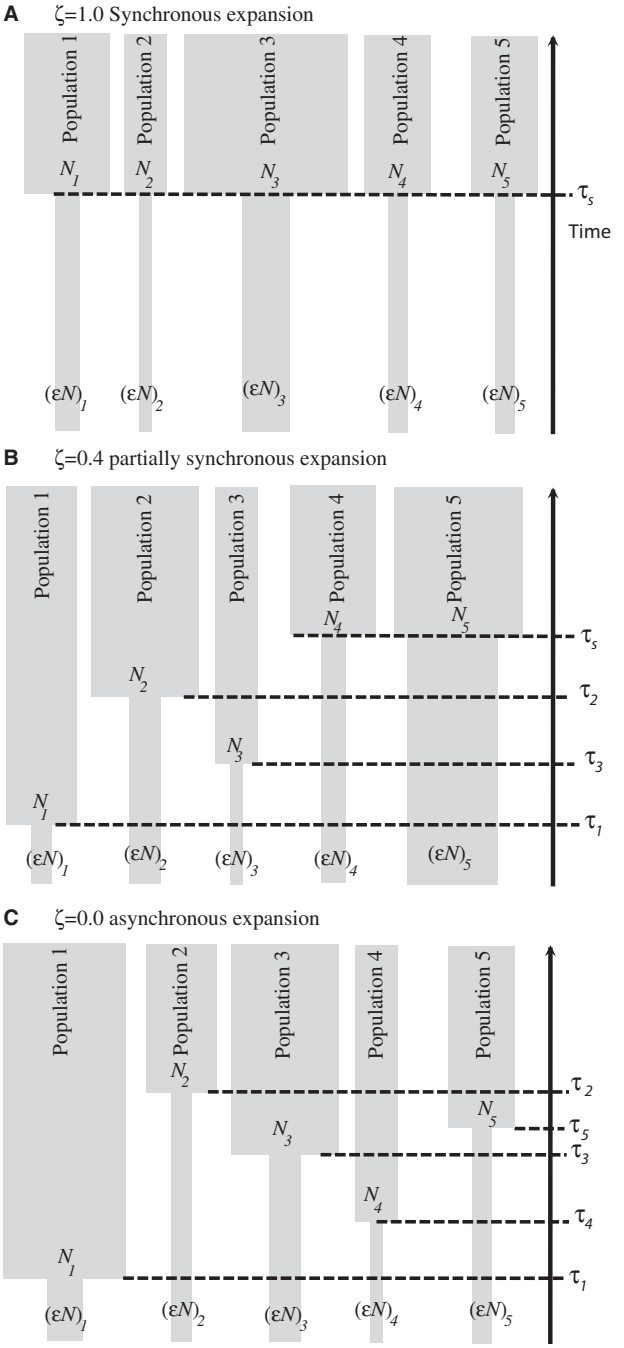

Fig. 1. Depiction of models of synchronous coexpansion (*A*; $\zeta = 1.0$), partially synchronous coexpansion (*B*; $\zeta = 0.4$), and asynchronous expansion (*C*; $\zeta = 0.0$) all involving five populations. Each population has a contemporary effective population size of N_i from prior $U(N_{min}, N_{max})$ and expands instantaneously from a population ε_i the current size chosen from expansion magnitudes $U(\varepsilon_{min}, \varepsilon_{max})$ at times τ_i or τ_s from the uniform prior of $U(\tau_{min}, \tau_{max})$.

the observed data (fig. 4), we estimated ζ to be 0.8 (95% quantiles = 0.3–1.0; fig. 5*A*), thereby corresponding to an estimate of 26 of 32 populations temporally coexpanding (95% quantiles = 10–32). The mode for the mean expansion time $E(\tau)$ was 54,321 (95% quantiles = 38,918–71,287) and of the time of the coexpansion τ_s was 35,225 (95% quantiles = 18,963–67,545; fig. 5*C* and *D*).

The hABC estimates of ζ, given five sets of 100 PODs each simulated from fixed values of ζ (0.0, 0.25, 0.5, 0.75, and 1.0), demonstrate that the method can detect the degree to which a community coexpanded at a single time, as well as estimate the expansion time and dispersion index of expansion times (fig. 6). For example, we simulated 100 PODs under the fixed value of $\zeta = 0.25$, thereby yielding 8 taxa expanding synchronously at time τ_s and 24 taxa expanding at random times τ_i, with the expansion times drawn from priors $\Pr(\tau_s)$ or $\Pr(\tau_i)$ of $U(1{,}000, 200{,}000)$ generations ago. Each estimate of ζ for each POD is based on 1,000 acceptances from a reference data table of 6.6 million simulations with ζ drawn from the discrete uniform hyperprior of $\Pr(\zeta) = [0.0, 1/32, \ldots, 31/32, 1.0]$. In this case, the reference table consists of 200,000 data sets simulated from each of the 33 possible values of the hyperprior ζ ranging from 0 to a total of 32 populations expanding congruently with parameters for each population drawn from species-specific priors. The cross-validation PODs and error estimates demonstrate that the hABC method can potentially obtain accurate estimates of ζ, $E(\tau)$, τ_s, and $\mathrm{Var}(\tau)/E(\tau)$.

As a heuristic approach to identify which 10–32 of the 32 avian populations coexpanded, we subsequently conducted an exploratory PCA. Specifically, we chose a cluster of 26 populations plotted on the first two components to identify a putative subset of populations that synchronously coexpanded (fig. 7*C*). We then assessed whether this subset of populations have data that are consistent with a history of temporal coexpansion by running an hABC analysis with a new reference table and observed summary statistic vector *D* regenerated from the subset of 26 populations corresponding to the PCA cluster. The estimates of ζ, $\mathrm{Var}(\tau)/E(\tau)$, and τ_s on this subset yield strong support for complete synchronous coexpansion (mode estimate of $\zeta = 0.95$, fig. 7) and agree with the estimate of coexpansion time derived from the previous analysis (mode estimate of 32 populations $\tau_s = 35{,}225$ vs. mode estimate of 26 populations $\tau_s = 33{,}465$, τ_s 95% quantiles = 26,059–50,961; fig. 7).

Discussion

The prevalence of shallow mtDNA genealogies found in many terrestrial and marine communities (Grant and Bowen 1998; Hewitt 2000) suggests that there may be widespread temporally and spatially shared responses of communities to Pleistocene environmental fluctuations such as sea level changes or glacial retreats. A statistical framework to test this hypothesis at the multispecies level will enable community-level inference for understudied groups and across whole biota. In contrast to previous applications of hABC to comparative phylogeographic data sets which sought to quantify support for models of vicariance or dispersal (Hickerson and Meyer 2008) or test for synchronous isolation (Huang et al. 2011), the method we present here enables quantifying the amount of temporal congruence in demographic coexpansion across species. Discerning whether species responded in concert or individualistically to historical changes in climate and landscape is important with respect to understanding the biogeographic dynamics of future climate changes, invasions, and extinction (Lavergne et al. 2010).

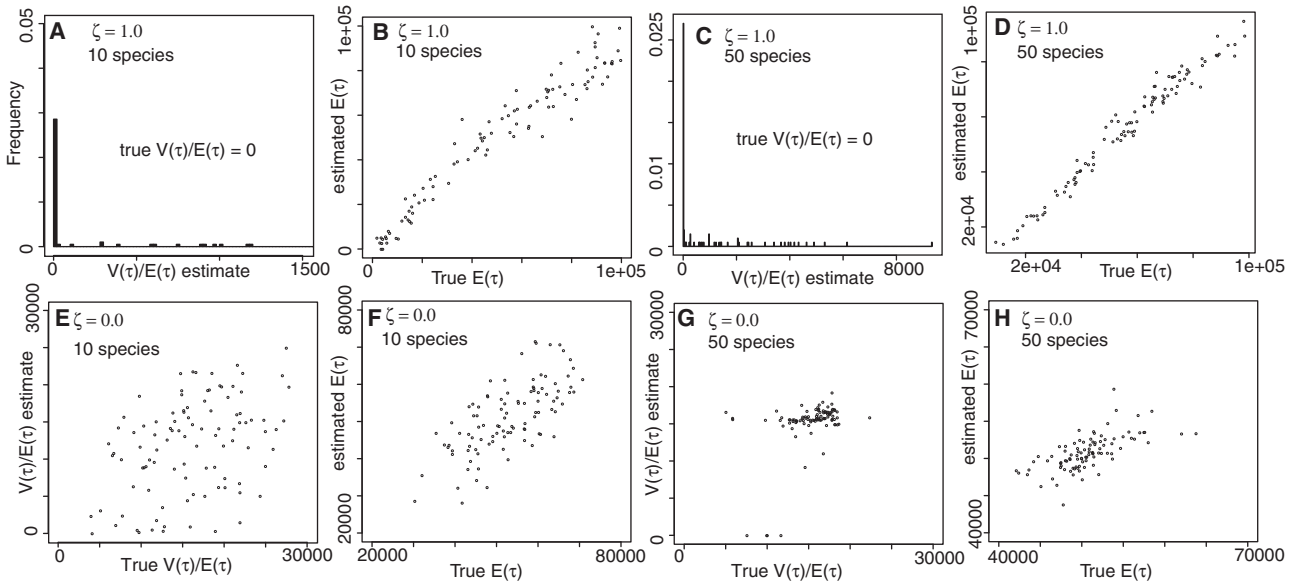

Fig. 2. Simulation validation of the ABC estimator of the mean and dispersion index of the expansion times across 10 (panels *A*, *B*, *E*, and *F*) or 50 (panels *C*, *D*, *G*, and *H*) populations ($E(\tau)$ and $\mathrm{Var}(\tau)/E(\tau)$, respectively). Each joint estimate was made on a POD with parameter values randomly drawn from the priors $\Pr(\tau) = U(1{,}000, 100{,}000)$ generations. Sets of PODs were drawn from histories with all populations expanding synchronously ($\zeta = 1.0$; panels *A*–*D*) and asynchronously ($\zeta = 0.0$; panels *E*–*H*). True value for $\mathrm{Var}(\tau)/E(\tau) = 0$ when $\zeta = 1.0$.

Table 2. Comparison of Different Numbers of Species and Expansion Time Priors Indicates that Using Greater Numbers of Species Lowers the Error in Detecting Simultaneous Expansion.

True Model (ζ)	Number of Species	τ_{max}	Error Rate (majority)	Error Rate (Bayes factor threshold > 3)	Error Rate (Bayes factor threshold > 10)
1	10	100,000	0.08	0.29	0.51
0	10	100,000	0.30	0.62	0.87
1	10	500,000	0.07	0.30	0.67
0	10	500,000	0.28	0.60	0.80
1	50	100,000	0.00	0.01	0.01
0	50	100,000	0.03	0.14	0.29
1	50	500,000	0.02	0.04	0.06
0	50	500,000	0.00	0.00	0.00

Note.—Error rates are calculated using either the majority of posterior samples or the Bayes Factor that compares the posterior weights of $\zeta = 0.0$ and $\zeta = 1.0$. Errors were determined from estimates of PODs generated under $\zeta = 0.0$ and $\zeta = 1.0$ for communities of 10 and 50 species and τ_{max} of 100,000 and 500,000 generations.

Although we focus on comparative demographic inference, interpretations from this method could focus on demographic histories of matrilineages or selective sweeps accompanied by hitchhiking (Gillespie 2001; Bazin et al. 2006).

This general method and downstream extensions will be able to be deployed on the wealth of mtDNA, single-copy nuclear DNA, and chloroplast spacer regions that have become available for animals, plants, protists, and fungi over the last quarter of a century for phylogeographic (Soltis et al. 2006) and mtDNA studies (Ratnasingham and Hebert 2007), as well as newly available ancient DNA data (Ramakrishnan and Hadly 2009; Ho and Gilbert 2010; Lorenzen et al. 2011). Even if only a single genealogical sample is available from each animal species, the opportunity to sample this locus within communities across wide taxonomic breadth enables researchers to make regional-scale inferences of how whole assemblages responded to the cyclical climate changes of the Pleistocene, all while accounting for the variability associated with using a single locus per species.

The demographic model involves synchronous and asynchronous expansion where each taxon consists of a single panmictic population that is expanding. Although heterogeneity in sample size, sequence length, mutation model, and generation time, are incorporated into the simulations and taxon-specific parameters such as mutation rate and effective population sizes are allowed to freely and independently vary across taxa, we suggest that when applying this method, taxa/populations should be chosen that are likely to be panmictic. If taxa are found to have population structure, the data set can be divided into populations and each subset included separately in the analysis, although one should be cautious in cases where there are an unknown number of unsampled ghost populations that could cause distortions in inference (Heller et al. 2013). As with other model-based methods, it is important to perform model-checking procedures, such as posterior predictive checks and cross-validation simulation experiments, to determine whether parameters can be accurately estimated given the empirical data (Gelman et al. 2004; Csilléry et al. 2010).

To examine the robustness of our model estimation procedure, we quantified the impact of the prior on posterior distributions, parameter inference, and model choice (Gelman et al. 2004). We found that the timescale of the prior had little influence on the model choice or parameter estimation but that error rates for model choice were lower with 50 species compared with 10 species (table 2). To determine goodness-of-fit and examine whether misspecified

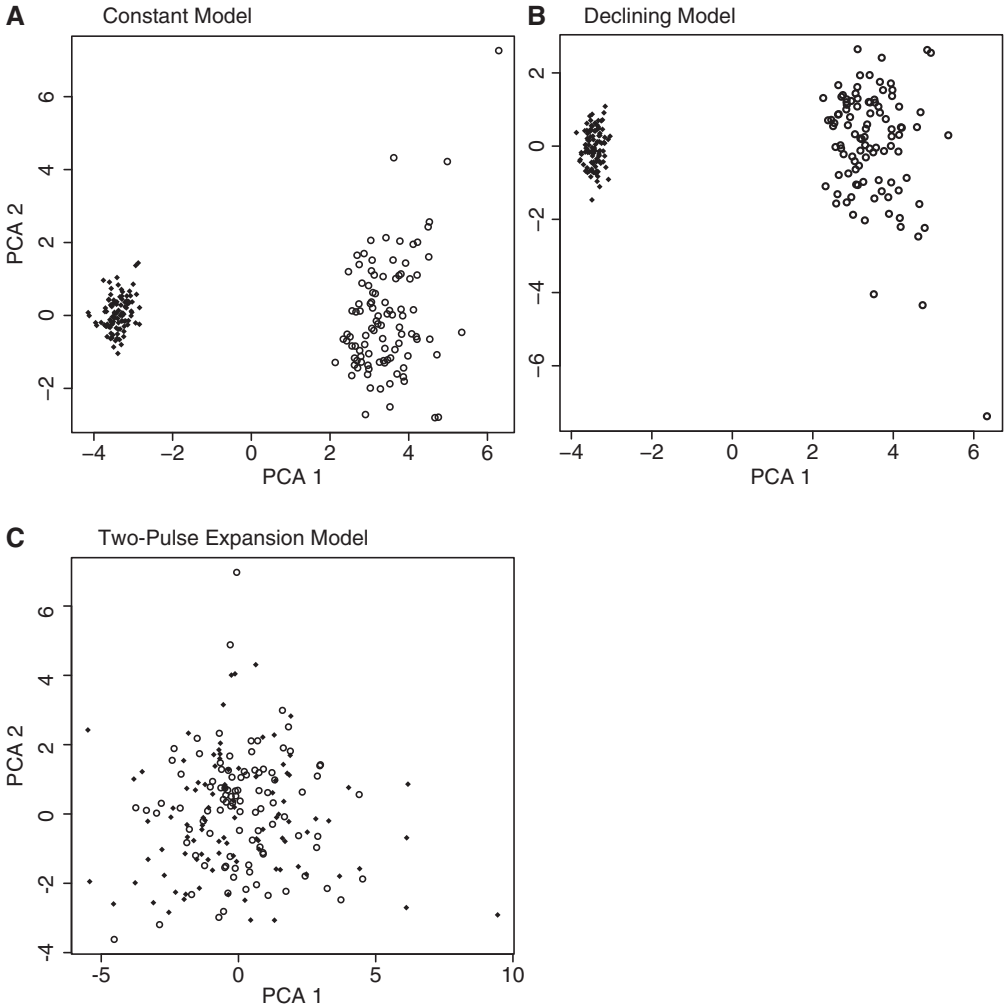

Fig. 3. Projection of the summary statistics calculated from 100 PODs from constant, declining, and two-pulse models. (*A*) Constant: 15 species expanding congruently at 1,000–100,000 generations before present and 35 species not expanding (◆). (*B*) Declining: 15 species expanding congruently at 1,000–100,000 generations before present and 35 species declining at 1,000–100,000 generations before present (◆). (*C*) Two-pulse model that included 35 species expanding congruently at 30,000 generations before present and 15 species expanding congruently at 90,000 generations before present (◆); and 100 samples from the hyperprior (o) of the expansion model where 15 species are expanding congruently at 1,000–100,000 generations before present and 35 are expanding at random times 1,000–100,000 generations before present onto the first two axes of a PCA.

models (i.e., when the actual model generating the data does not match the model used for inference) could be detected, we performed graphical model checking on PODs from constant, declining, and two-pulse models. We found that a PCA comparing summary statistics calculated from PODs generated from these alternative models and summary statistics calculated from data generated from random draws from the prior model used for inference (i.e., our expansion model) was effective for identifying cases of poor model fit for the constant and declining population models (fig. 3*A* and *B*) but not the two-pulse model (fig. 3*C*). However, in cases where the data are generated under a two-pulse model, one can first use our method to detect asynchronous coexpansion and subsequently explore running the framework on subsets of the data to identify multiple pulses of coexpansion (fig. 7).

To minimize bias introduced by violations of the assumption of panmictic populations (Navascués and Emerson 2009; Heller et al. 2013), we demonstrate the method on an avian assemblage from Australia, where species typically have high dispersal and consequently simplified patterns of population structure. In Australia, many common widespread species across the continent may have had assemblage-wide bottlenecks in response to widespread climatic events in the late Pleistocene (Barnosky et al. 2004), and our method suggests that a subset of populations potentially coexpanded 26,000–51,000 generations ago, coinciding with Marine Isotope Stage 3, a highly variable climatic period (Siddall and Rohling 2008).

Although hABC enabled the estimation of the number of species in the Australian data set that share congruent demographic dynamics, the identification of individual species by PCA shown here is heuristic, and there are likely to be other subsets that may also yield high posterior probability of synchronous coexpansion. Beyond such exploration with PCA, one could examine parameter estimates from individual species analysis using Bayesian skyline plots or mismatch distribution methods (Rogers and Harpending 1992; Kuhner et al. 1998; Schneider and Excoffier 1999; Ray et al. 2003; Excoffier 2004; Ho and Shapiro 2011). Whatever method is

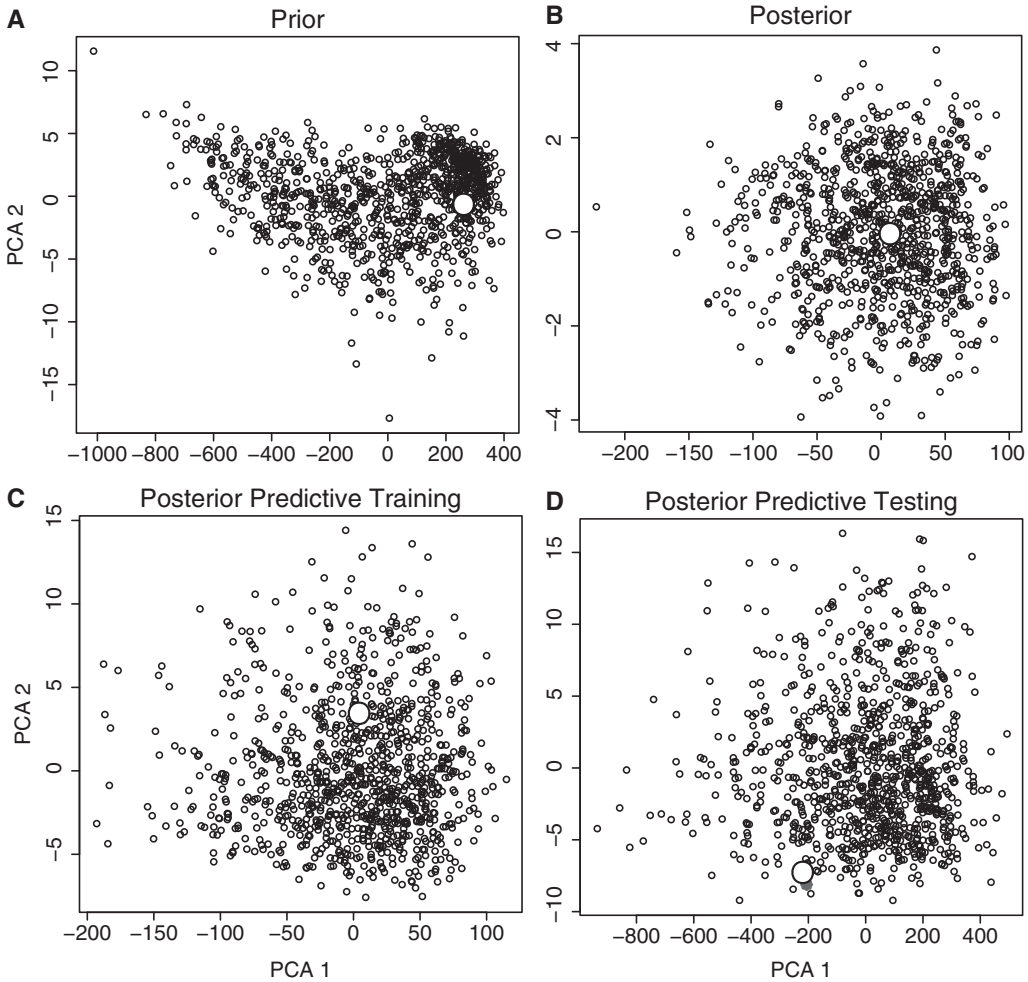

Fig. 4. Projection of the summary statistics calculated from 1,000 samples from the prior (*A*), posterior (*B*), and posterior predictive distributions (*C* and *D*) onto the first two axes of a PCA. The open circle is calculated from the observed Australian data collected from 32 avian populations. Panels (*A*–*C*) use the 16 original summary statistics used for the parameter estimates (mean, variance, skewness, and kurtosis of the number of haplotypes, haplotypic diversity, nucleotide diversity, and Tajima's *D*) and panel (*D*) uses an alternative set of 8 summary statistics (mean, variance, skewness, and kurtosis of the number of segregating sites and Fu's *F*). In panels (*C*) and (*D*), the 1,000 sets of accepted parameter values conditioned on the observed data to approximate the posterior were subsequently used to resimulate the data 1,000 times to obtain the posterior predictive distributions (Gelman et al. 2004; Cornuet et al. 2010).

used to identify the individual species, the subset of congruent species can then be verified with an hABC analysis on the putative subset and with resulting parameter estimates of $\zeta \approx 1.0$ and $\mathrm{Var}(\tau)/E(\tau) \approx 0$ providing confirmation that the subset of species expanded synchronously, although competing models with alternate subsets may be unidentifiable given single locus mtDNA data.

We model an assemblage history that involves one synchronous expansion plus a group of asynchronous expansions, although future approaches could incorporate more than one synchronous expansion pulse and/or pulses or contraction pulses. The use of a more realistic model of population growth such as an exponential model may also improve inference, although our instantaneous model is likely to capture the effects of an exponential model (Rogers and Harpending 1992). Greater model complexity may also be accommodated by the incorporation of high throughput population genomic data obtained from next-generation sequencing technology.

Conclusion

We present a new hABC method for demographic inference and testing community response to climate change. This framework enables the estimation of the proportion of the present-day community that expanded in a single pulse in the past and will further aid in fitting predictive species distributional models of shared demographic and distributional changes (Lorenzen et al. 2011; Prunier et al. 2012). We test this method on an empirical data set of avian populations from Australia and demonstrate our ability to detect and estimate the timing of the expansion pulse. This method could help test macroecological hypotheses of community assembly (Weiher et al. 2011) and identify different responses by different ecological guilds or interacting species (Emerson and Gillespie 2008; Mikheyev et al. 2008; Stone et al. 2012). Overall, this approach can be extended to enable detecting forces underlying regional biodiversity patterns and provide a greater understanding of how changes in historical

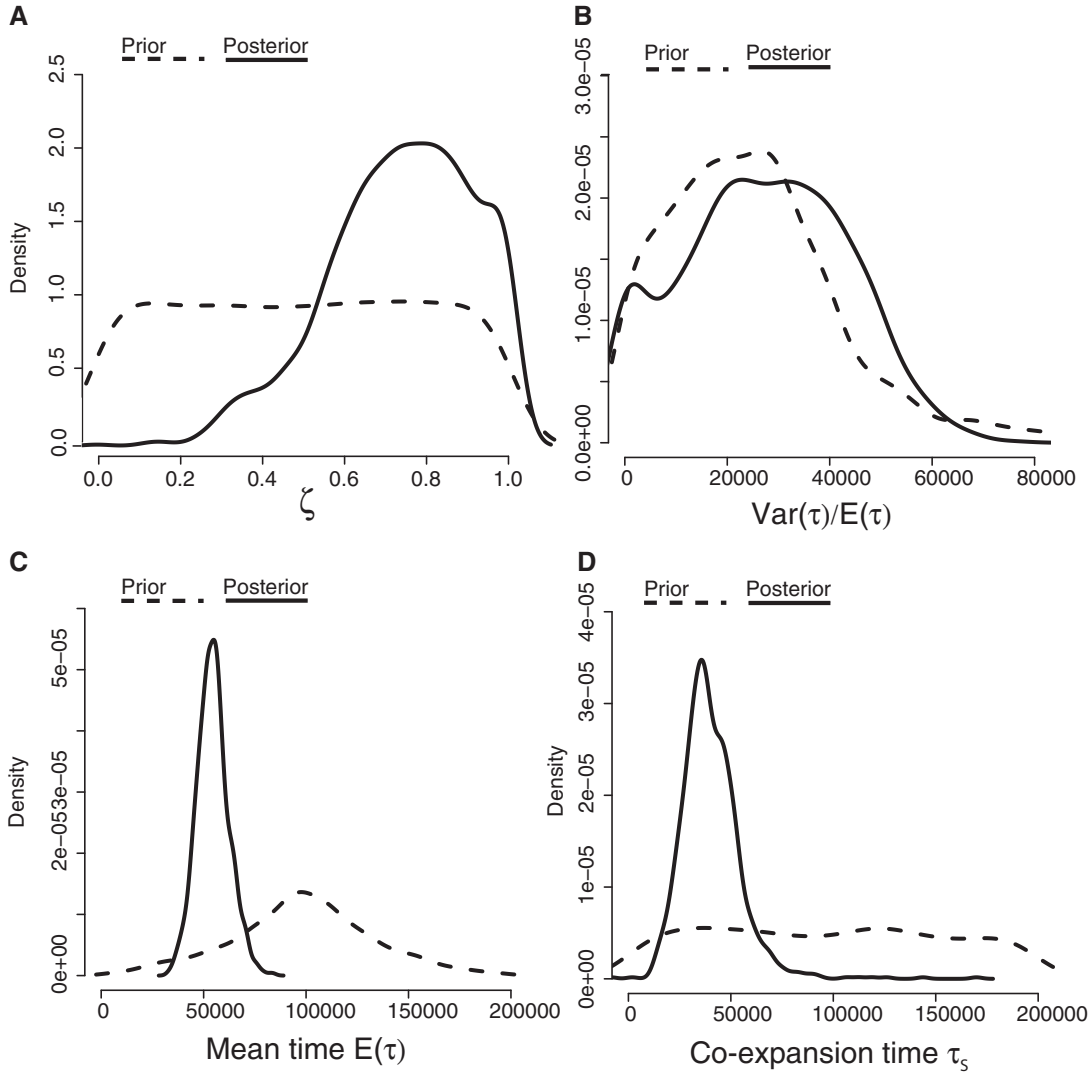

Fig. 5. Posterior estimates of model index of community congruence hyperparameter ζ (panel *A*) and parameter summaries time dispersion index $\mathrm{Var}(\tau)/E(\tau)$ (panel *B*), total mean expansion time $E(\tau)$, and coexpansion time τ_s (panels *C* and *D*) conditional on data from 32 Australian avian populations. The posterior of ζ with a mode between 0.7 and 0.8 and $\mathrm{Var}(\tau)/E(\tau)$ »0.0 indicates a mixed history that includes populations that expanded asynchronously and a synchronous pulse of coexpansion in a subset of the populations.

environmental features differentially affected species distributions (Hickerson et al. 2010; Emerson et al. 2011).

Materials and Methods

The hABC Model

Similar to the hierarchical Bayesian formulation of Ilves et al. (2010), which was used to estimate the proportion of taxon pairs that arose by way of colonization rather than "soft vicariance," our objective here is to estimate the community congruence hyperparameter ζ, the proportion of n taxa that coexpanded in a single synchronous cluster. A flow chart of our method is outlined in table 1. The hierarchical structure of the model is such that the data D is conditional on species-specific parameters ϕ and τ, whereby the expansion time parameters τ_i for $i = 1, 2, \ldots, n$ species are conditional on ζ, whereas the other species-specific parameters ϕ_i (μ_i, N_i, and ε_i; mutation rates, current effective population sizes, and expansion magnitudes, respectively) are assigned to $i = 1, 2, \ldots, n$ species independent of ζ. Under this scheme, the posterior distributions of the parameters given the data are $\pi(\zeta,\tau \mid D) \propto P(D \mid \zeta,\tau)\pi(\tau \mid \zeta)\pi(\zeta)$, and $\pi(\phi \mid D) \propto P(D \mid \phi)$ with the joint hierarchical posterior distribution of the parameters given the data being $\pi(\zeta, \tau, \phi \mid D) \propto P(D \mid \zeta,\tau,\phi) = P(D \mid \phi,\tau)\pi(\phi)\pi(\tau \mid \zeta)\pi(\zeta)$. The community congruence hyperparameter ζ is then modeled, such that it can take $k = n + 1$ possible values $\zeta\{0.0, 1/n, 2/n, \ldots, n/n\}$ that are assigned prior probabilities drawn from a discrete uniform distribution $\pi(\zeta)$. In this case, each model ζ_k within the vector $\{\zeta_0, \ldots, \zeta_1\}$ has equal prior probability, and the expansion time parameters τ_i are assigned to each of the n species conditional on ζ, $\pi(\tau \mid \zeta)$.

Under the ABC scheme, this hierarchical Bayesian mixture model is used to simulate the data D in three steps: 1) generating the model ζ_k from $\pi(\zeta)$ with each ζ value having equal probability $P(\zeta_{0.0}), \ldots, P(\zeta_{1.0})$; 2) generating the parameter vectors ϕ_k and τ_k from $\pi(\phi)$ and $\pi(\tau \mid \zeta)\pi(\zeta)$; and 3) generating the data D from $P(D \mid \phi_k,\tau_k)$. By conditioning on the

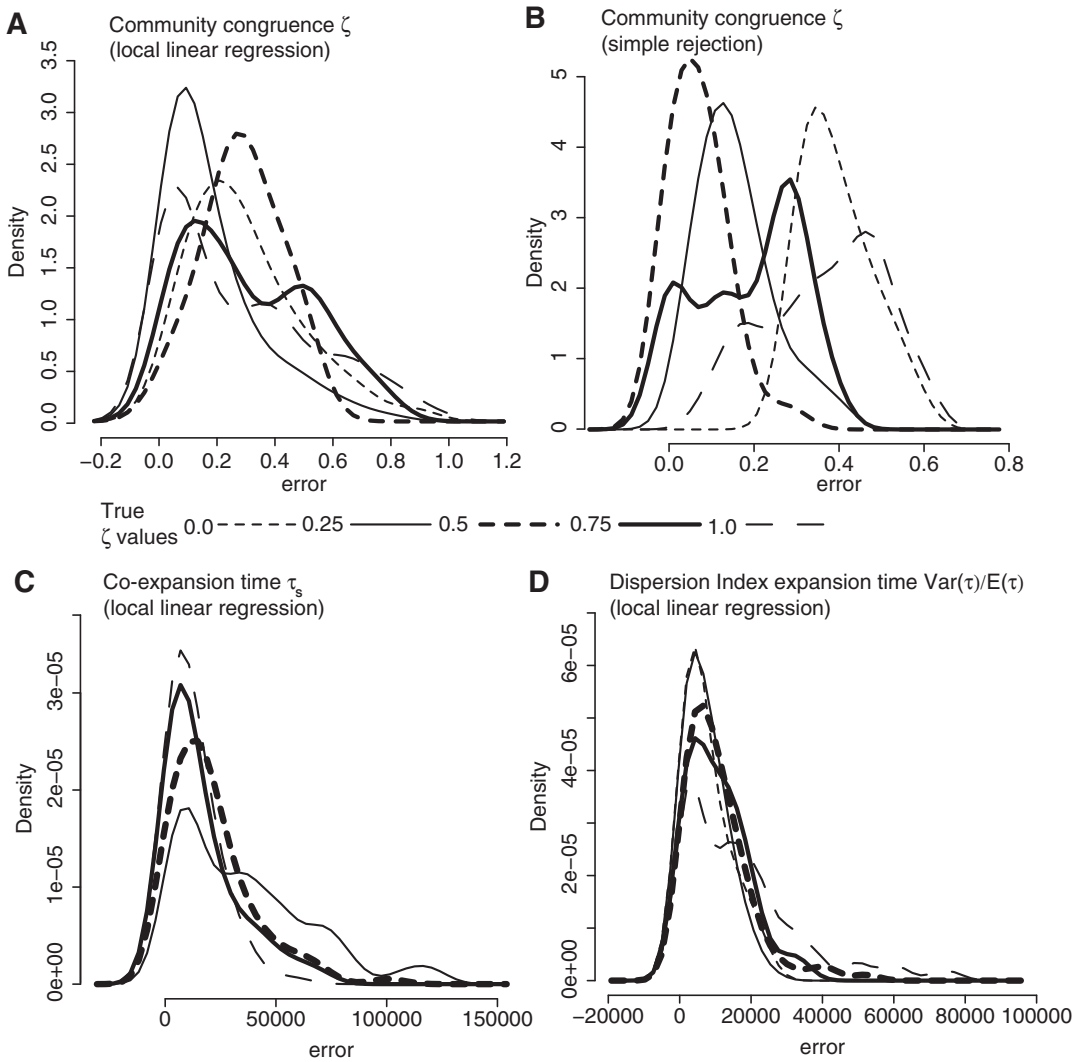

Fig. 6. Error estimates from simulation validation of the hABC estimator of model index of community congruence ζ using local linear regression (panel *A*), simple rejection (panel *B*), τ_s coexpansion time (panel *C*), and $\mathrm{Var}(\tau)/E(\tau)$ time dispersion index of all expansion times (panel *D*). A set of 100 PODs was generated for five different levels of community congruence; true $\zeta = 1.0$ where all species expanded in a synchronous pulse, true $\zeta = 0.0$ where all species expanded randomly, and intermediate levels of synchronous and asynchronous expansion $\zeta = 0.25, 0.5, 0.75$. Each set of 100 PODs was based on the sampling and priors of 32 Australian bird populations with expansion time randomly drawn from the prior $\Pr(\tau) = U(1{,}000, 200{,}000)$.

data *D*, this hierarchical ABC model then yields the approximate posterior probabilities of the focal hyperparameter ζ and parameter summaries $E(\tau)$,τ_s, and $\mathrm{Var}(\tau)/E(\tau)$, while allowing other taxon-specific parameters to be drawn independently for each taxon. For an analysis involving *n* taxa, these taxon-specific parameters ϕ include the *n* current effective population sizes $\{N_1, \ldots, N_n\}$ drawn from a uniform prior $P(N_k) = \sim U(N_{\min}, N_{\max})$, the *n* expansion magnitudes $\{\varepsilon_1, \ldots, \varepsilon_n\}$ drawn from a uniform prior $P(\varepsilon_k) = \sim U(\varepsilon_{\min}, \varepsilon_{\max})$, and the *n* mutation rates $\{\mu_1, \ldots, \mu_n\}$ that are in turn drawn from a uniform prior $P(\mu_k) = \sim U(\mu_{\min}, \mu_{\max})_\mu$ with each locus-specific mean chosen according to the data and estimates based on the literature (table 3, Nabholz et al. 2008). With regard to each taxon's expansion times $\tau_a = \{\tau_1, \ldots, \tau_{\,|(1-\zeta)\eta\,|}\}$, the $||\,(1-\zeta)n\,||$, independently expanding taxa are each independently and randomly assigned an expansion time from the uniform prior $P(\tau_a) = \sim U(\tau_{\min}, \tau_{\max})$, whereas the $||\,n\zeta\,||$ taxa synchronously coexpanding are assigned a single τ_s that is randomly drawn from the same uniform $\sim U(\tau_{\min}, \tau_{\max})$ ($||\,..\,||$ notation specifies the integer nearest the number within).

Importantly, all *n* taxa experience an instantaneous expansion of magnitude ε_k at τ_s or $\tau_a = \{\tau_1, \ldots, \tau_{\,|(1-\zeta)\eta\,|}\}$ generations in the past to reach their respective effective size N_k, such that each taxon's effective population size at time τ_s or times $\tau_a = \{\tau_1, \ldots, \tau_{\,||(1-\zeta)\eta\,||}\}$ is $(\varepsilon N)_k < N_k$ as depicted in figure 1 under the three different hypothetical scenarios of $\zeta = 1.0$, 0.4, and 0.0, respectively. As a practical way to estimate the hyperparameter model indicators for $\zeta_{0.0}, \ldots, \zeta_{1.0}$ from the data, we choose the hyperprior $\Pr(\zeta_k)$ to be a simple discrete uniform prior $P(\zeta_k) \equiv 1/k$ to favor all models equally and where the posterior is proportional to the likelihood $\pi(\zeta,\tau,\phi \mid D) \propto P(D \mid \zeta,\tau,\phi) = P(D \mid \phi,\tau)\pi(\phi)\pi(\tau \mid \zeta)\pi(\zeta)$. In this case, there is one model ζ_k for each number of possible taxa coexpanding, such that if there are 100 taxa, then there are 101 models of ζ_k.

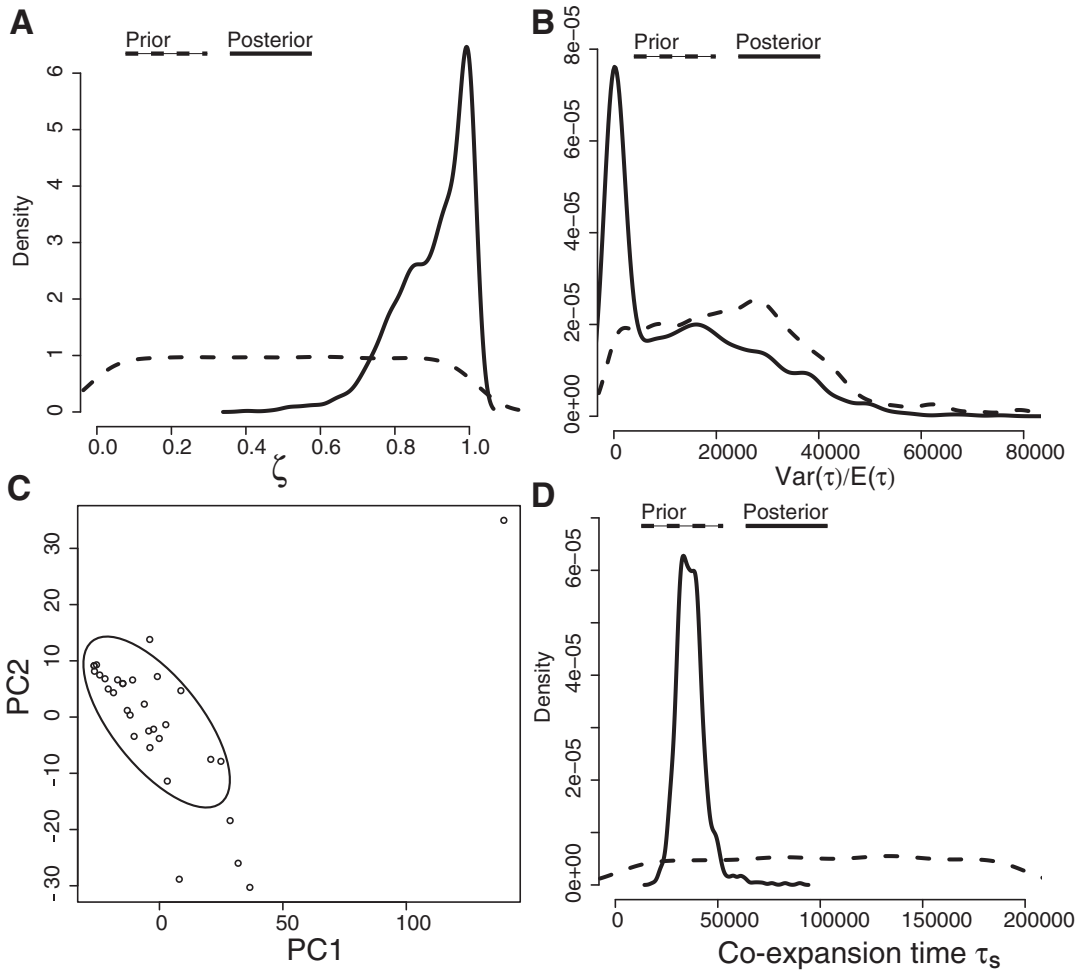

Fig. 7. Posterior estimates of hyperparameter ζ and parameter summaries τ_s and $\mathrm{Var}(\tau)/E(\tau)$ conditional on data from the subset of 26 Australian bird populations that were identified by the PCA to have synchronously coexpanded. The ABC posterior distributions indicate that these 26 populations expanded synchronously (panels *A* and *B*) approximately 33,000 years ago (panel *D*) during Marine Isotope Stage 3. Panel *C* depicts the first two principal components calculated from the 16 summary statistics calculated from the 32 Australian avian populations. The tightest cluster of 26 species was subsequently tested as a putative subset of species that synchronously coexpanded.

Using this hABC approach, we compress an observed multitaxon data set into a summary statistic vector (D^*) and condition on D^* to obtain $\pi(\zeta,\tau,\phi \mid D^*)$, the approximate joint posterior probabilities of the expansion times τ_k (summarized by posterior estimates of $E(\tau)$, $\mathrm{Var}(\tau)/E(\tau)$) and ζ, the proportion of the taxon assemblage that expanded synchronously. In this case, the data consist of a sample of multiple alleles from a single locus from each of n taxa assumed to be panmictic.

Multitaxa coalescent simulations were performed using a python script that combined single taxa simulations and summary statistics from Bayes Serial Simcoal (Anderson et al. 2005) into a single file and calculates the multispecies summary statistics. Python scripts for the simulations are available at https://github.com/UH-Bioinformatics/hBayeSSC (last accessed June 16, 2014).

Summary Statistics

Following our objectives of quantifying temporal patterns of coexpansion, we construct multispecies summary statistics that are based on four summary statistics that are known to be correlated with demographic expansion, including number of haplotypes, haplotypic diversity, nucleotide diversity, and Tajima's *D*. To ensure that the summary statistic vector *D*, used for the ABC procedure, is independent of the ordering of the data configuration, and to reduce the dimensionality in *D*, we use the first four sample moments of these four summary statistics calculated on each of the four summary statistics across the n taxa. Specifically, we use the mean, variance, skewness, and kurtosis of each of these four summary statistics thereby yielding a vector *D* consisting of 16 order-independent multitaxa summary statistics. To initially explore the statistical behavior of the 16 components of D^*, we simulated data from 50 taxa and a prior approximately $U(1{,}000, 500{,}000)$ for τ under models $\zeta = 0.0$ and $\zeta = 1.0$. These conditions demonstrated the summary statistics to have differing distributions under these two models, a favorable condition for summary statistic selection in ABC (supplementary figs. S1 and S2, Supplementary Material online; Marin et al. 2012).

hABC Estimation

For obtaining the ABC joint posterior of ζ_k and the summaries of the n expansion times ($E(\tau)$, τ_s, $\mathrm{Var}(\tau)/E(\tau)$), we use rejection sampling to identify the 10,000 closest Euclidean distances between D^* and D_i calculated from 2.0×10^6 random draws from each of the models $P(\zeta_k) = U(\zeta_0, \ldots \zeta_1)$ with equal prior probability. After a second step of rejection sampling to identify the 1,000 closest Euclidean distances between D^* and D_i, we fit a local linear regression model to these posterior draws and applied this model to these remaining 1,000 acceptances to obtain an adjusted estimate of the joint posterior probability of $E(\tau)$, τ_s, $\mathrm{Var}(\tau)/E(\tau)$, and ζ_k using functions available from the R package abc.r (Csilléry et al. 2012). The local linear adjustment leads to improved estimation over estimates obtained by simple rejection sampling (fig. 6*A* and *B*). For ζ_k, we then back-transform the adjusted estimates to the closest value contained in the discrete uniform prior $P(\zeta_k)$ due to local linear regression leading to adjusted values of ζ that are not contained in the prior. For example, if the prior contains 0, 1/50 = 0.02, 2/50 = 0.04, . . . , 50/50 = 1, and the parameter value sampled in the posterior after the regression adjustment is 0.025, this value is set to 0.02 (1/50).

Simulation Cross-Validation of ABC Estimation

The development of novel ABC methods require careful examination of bias and accuracy, and especially in the case of using ABC for model estimation, these developments must go hand in hand with validating simulation experiments (Cook et al. 2006; Bertorelle et al. 2010; Cornuet et al. 2010; Csilléry et al. 2010; Robert et al. 2011). To this end, we first simulated 10-species and 50-species models, while varying the priors for expansion times using a $P(\tau_{\min}, \tau_{\max})$ of $U(1{,}000, 100{,}000)$ and $U(1{,}000, 500{,}000)$ generations. The sample configuration of the 10 and 50 species consisted of 40 sequences of 800 bp of single-locus mtDNA data and a fixed locus mutation rate of 8×10^{-5}.

Initially, the prior distribution for ζ, $P(\zeta)$, only considered the two extreme values of $\zeta = 0.0$ and $\zeta = 1.0$, by simulating 500,000 prior draws from these two models. We then independently simulated 100 PODs (Bertorelle et al. 2010) from each of these two models and then used the ABC procedure to yield the 1,000 closest Euclidean distances between each PODs' D^* and D_i calculated using the data generated from each of the 500,000 prior draws and used this to approximate the posterior probabilities of $\zeta = 0.0$ and $\zeta = 1.0$ for each estimate. Additionally, we used the Bayes Factor $P(\zeta = 0.0 \mid D^*)/P(\zeta = 1.0 \mid D^*) \div P(\zeta = 0.0)/P(\zeta = 1.0)$ or $P(\zeta = 1.0 \mid D^*)/P(\zeta = 0.0 \mid D^*) \div P(\zeta = 1.0)/P(\zeta = 0.0)$ (Kass and Raftery 1995) and the associated Jeffrey's scale (Jeffreys 1961) to gauge one's support for either history ($\zeta = 0.0$ or $\zeta = 1.0$). As an additional assessment, we estimated $\mathrm{Var}(\tau)/E(\tau)$ and $E(\tau)$ from 100 PODs conditional on $\zeta = 0.0$ and $\zeta = 1.0$, and either of the two numbers of populations (10 and 50) and priors for $P(\tau) = U(1{,}000, 100{,}000)$ and $U(1{,}000, 500{,}000)$.

Detecting Poor Model Fit with Graphical Checks

We used the samples from the prior to check that our model priors could generate summary statistics similar to those calculated from PODs generated from alternative models (Cornuet et al. 2010). We performed a PCA using the "prcomp" function in R (R Development Core Team 2008) on the 16 multitaxa summary statistics from each of 100 PODs simulated from our hyperprior conditional on $\zeta = 0.4$ given a history of 15 coexpanding taxa at 1,000–100,000 generations before present and 35 taxa expanding at independent times and 100 PODs generated from three alternative model priors: 1) constant—15 taxa coexpanding at 1,000–100,000 generations before present and 35 taxa not expanding; 2) declining—15 taxa coexpanding at 1,000–100,000 generations before present and 35 taxa declining at 1,000–100,000 generations before present; and 3) two pulse—35 taxa coexpanding at 30,000 generations before present and 15 taxa coexpanding at 90,000 generations before present. For the graphical checks, we specifically focused on PC1 and PC2 for comparison.

Pleistocene Expansion Times in Australian Avian Populations

We apply our method to a set of 32 avian populations distributed from across the Australian continent, most of which have been previously published and analyzed (table 3). Each population consists of a monophyletic cluster within which the population can be considered to be demographically panmictic. Each population putatively expanded from a single refugium during the Pleistocene, and our method attempts to discern if any of them coexpanded temporally without making any spatial assumptions or inferences. Following the general hierarchical ABC procedure, we simulated from the discrete uniform prior of $P(\zeta_k) = U(\zeta_0, \ldots, \zeta_1)$, with 200,000 simulations from each of the 33 coexpansion models, which ranged from 0 to 32 species coexpanding at time τ_s (total 6.6 million simulations). Following this hABC structure, τ_s and $\tau_a = \{\tau_1, \ldots, \tau_{|(1-\zeta)\eta|}\}$ are randomly drawn from the uniform prior $P(\tau) = U(1{,}000, 200{,}000)$. Likewise, the effective size N_i of each of the 32 contemporary populations is independently drawn from a species-specific uniform prior $U(\mathrm{Ne}_{i_\min}, \mathrm{Ne}_{i_\max})$ that instantaneously expands from an effective population ε_i its size at time τ_s or time $\tau_a = \{\tau_1, \ldots, \tau_{|(1-\zeta)\eta|}\}$. Mutation rates were drawn from a uniform prior distribution of $U(\mu_{\min}, \mu_{\max})$ that were based on previously reported locus and taxon-specific rates (table 3; Nabholz et al. 2008). After 6.6×10^6 simulated random draws from the prior, the ABC filter of 1,000 closest Euclidian distances between observed D^*, and the 6.6×10^6 simulated values of D_i are retained. Subsequently, we fit a local linear regression model to these posterior draws and apply this model to obtain an adjusted estimate of the joint posterior probability of ζ_k, $E(\tau)$, τ_s, and $\mathrm{Var}(\tau)/E(\tau)$. To further ascertain our ability to estimate the number of coexpanding populations, we estimated ζ, $E(\tau)$, τ_s, and $\mathrm{Var}(\tau)/E(\tau)$ from sets of 100 PODs that were drawn from $\zeta = 0.0$, $\zeta = 0.25$, $\zeta = 0.5$,

Table 3. Summary and Sample Statistics for 32 Australian Birds Used in the hABC Analysis.

Taxa	No. Samples	No. Base Pairs	Mutation Rate Prior (%/My)	Nucleotide Diversity	No. Haplotypes	Haplotypic Diversity	Tajima's *D*	*P*-Value Tajima's *D*	Reference
Gymnorhina tibicen clade A	1,153	590	13–17	0.008110603	46	0.875974	−0.0657334	0.557	Baker et al. (2000)
Pomatostomus temporalis clade A[a]	84	399	13–17	0.010191312	21	0.849971	−0.261851	0.477	Edwards (1993)
Barnardius zonarius clade A[a]	11	1,036	0.6–1.0	0.000491502	3	0.472727	−0.778153	0.311	Joseph and Wilke (2006)
B. zonarius clade B[a]	70	1,036	0.6–1.0	0.00366875	13	0.701863	−1.44079	0.056	Joseph and Wilke (2006)
Malurus splendens clade A[a]	63	985	2.8–4.8	0.009792042	28	0.856631	−0.304745	0.461	Kearns et al. (2009)
Entomyzon cyanotis clade A	6	1,032	2.8–4.8	0.055232558	4	0.8	−1.52057	0.001	Toon et al. (2010)
Platycercus elegans clade A[a]	82	935	0.6–1.0	0.00308655	28	0.841012	−1.50608	0.054	Joseph et al. (2008)
M. lambertii clade A	63	969	2.8–4.8	0.007084969	39	0.950333	−1.26882	0.084	McLean et al. (2012)
Eopsaltria australis clade A[a]	53	1,002	2.8–4.8	0.001715967	18	0.716981	−2.28436	0	Pavlova et al. (2013)
E. australis clade C[a]	44	1,002	2.8–4.8	0.002325159	18	0.900634	−1.15957	0.1	Pavlova et al. (2013)
Cracticus quoyi clade A[a]	30	630	2.8–4.8	0.006871009	8	0.770115	1.05393	0.879	Kearns et al. (2011)
Gliciphila melanops clade A[a]	27	865	2.8–4.8	0.005643745	18	0.94302	−1.3094	0.092	Dolman and Joseph (2012)
Cinclosoma cinnamomeum clade A[a]	18	884	2.8–4.8	0.006750362	13	0.901961	−1.74447	0.025	Toon et al. (2012)
Ci. cinnamomeum clade B[a]	9	884	2.8–4.8	0.006473102	8	0.972222	−0.414598	0.399	Toon et al. (2012)
M. leucopterus clade A[a]	34	1,007	2.8–4.8	0.00448793	18	0.932264	−1.47934	0.051	McLean et al. (2012)
Ptilonorhynchus violaceus clade A[a]	16	1,002	2.8–4.8	0.002170229	9	0.925	−0.750111	0.235	Nicholls and Austin (2005)
Pt. violaceus clade B	78	1,002	2.8–4.8	0.004785263	47	0.968698	−1.98111	0.003	Nicholls and Austin (2005)
M. melanocephalus clade A[a]	14	467	2.8–4.8	0.000611808	3	0.274725	−1.48074	0.138	Lee and Edwards (2008)
M. melanocephalus clade B[a]	15	467	2.8–4.8	0.006322015	7	0.838095	−0.490602	0.379	Lee and Edwards (2008)
Biziura lobata clade A[a]	16	373	13–17	0.003024194	3	0.575	0.708323	0.783	Guay et al. (2010)
Bi. lobata clade B[a]	135	373	13–17	0.002248703	8	0.413709	−1.41781	0.047	Guay et al. (2010)
C. nigrogularis clade A[a]	55	867	2.8–4.8	0.001648162	20	0.642424	−2.42327	0	Kearns et al. (2010)
Chalcites minutillus clade A[a]	48	1,497	2.0–4.0	0.004713327	27	0.953014	−1.10275	0.147	Joseph et al. (2011)
Anas gracilis clade A	72	609	13–17	0.013997121	52	0.986307	−1.05892	0.149	Joseph et al. (2009)
Stipiturus malachurus clade A[a]	75	281	2.8–4.8	0.012100259	13	0.824144	−0.711011	0.254	Donnellan et al. (2009)
S. malachurus clade B	4	281	2.8–4.8	0.044483986	2	0.5	−0.858583	0.181	Donnellan et al. (2009)
P. halli clade A[a]	20	403	13–17	0.002638109	6	0.573684	−1.18349	0.115	Lee and Edwards (2008)
P. halli clade B[a]	5	403	13–17	0.004472057	5	1	−0.410175	0.526	Lee and Edwards (2008)
Pezoporus wallicus clade B[a]	22	849	0.6–1.0	0.000504492	4	0.547619	−0.849505	0.268	Murphy et al. (2010)
Ch. basalis clade A[a]	21	1,689	2.0–4.0	0.001438751	14	0.938095	−1.74716	0.022	Joseph et al. (2002)
Ci. punctatum clade A[a]	11	884	2.8–4.8	0.002838338	8	0.890909	−1.90258	0.012	Toon et al. (2012)
Phylidonyris novaehollandiae clade B[a]	20	1,033	2.8–4.8	0.00462628	9	0.873684	−1.44806	0.043	Dolman and Joseph (2012)

[a]Species were identified by the PCA as a possible coexpansion subset. Locus and taxon-specific avian mutation rate prior based on Nabholz et al. (2008).

$\zeta = 0.75$, and $\zeta = 1.0$ and calculated the error as $|(\text{true value} - \text{estimated value})|$.

Model Checking of the Prior and Posterior Predictive Distributions

As a goodness-of-fit check to ascertain whether the model and our chosen priors can produce the main features of the observed data and to check for prior sampling efficiency (Hickerson et al. 2014), we deployed a PCA on prior and posterior samples of the 16 summary statistics. Initially, we used the first two principal components of the summary statistics calculated from 1,000 random draws from the simulated prior distribution and 1,000 samples from the hABC posterior to compare with these first two components from the 32 avian population samples (D^*). Second, we sample from the posterior predictive distribution (Gelman et al. 2004) by simulating 1,000 data sets using parameters from the 1,000 posterior samples. To this end, we used the same 16 summary statistics, the mean, variance, skewness, and kurtosis of the number of haplotypes, haplotypic diversity, nucleotide diversity, and Tajima's D ("training" set) as well as an alternative set ("testing" set) that included the mean, variance, skewness, and kurtosis of the number of segregating sites and Fu's F and again projected the first two principal components of the corresponding observed summary statistics in both cases.

Selecting Taxa with Shared History

After the posterior estimate of ζ was obtained, given the observed D^* from 32 avian population samples, we used a PCA on the summary statistics to aide in selecting and testing which ($|\eta\zeta|$) populations coexpanded synchronously. Specifically, we calculated the first two principal components of the summary statistics calculated from 32 avian population samples. After recalculating D^* on a subset of populations that are closely clustered in the PCA, we then perform an hABC analysis to determine if this subset of populations plausibly coexpanded synchronously.

Supplementary Material

Supplementary figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank the Australian Center for Ecological Synthesis working group on Reconstructing Effects of Past Climate Change, including A. Cooper, J. Soubrier, L. Joseph, and V. Thompson. They also thank the editor and two anonymous reviewers for their insightful comments and suggestions that helped us improve the manuscript. This work was supported by a Marie Curie International Incoming Fellowship within the 7th European Community Framework Programme, National Science Foundation, Division of Ocean Sciences (Award 1260169 to R.J. Toonen and B. Bowen) to Y.L.C. and by the National Science Foundation, Division of Environmental Biology (Awards 1253710 and 1343578) to M.J.H. Computational support was provided by the National Center for Research Resources (5P20RR016467-11) and the National Institute of General Medical Sciences (8 P20 GM 103466-11) from the National Institutes of Health. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575.

References

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1733–1734.

Baker A, Mather P, Hughes J. 2000. Population genetic structure of Australian magpies: evidence for regional differences in juvenile dispersal behaviour. *Heredity* 85:167–176.

Barber BR, Klicka J. 2010. Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. *Proc Biol Sci.* 277: 2675–2681.

Barnosky A, Koch P, Feranec R. 2004. Assessing the causes of late Pleistocene extinctions on the continents. *Science* 306:70–75.

Bazin E, Dawson KJ, Beaumont MA. 2010. Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics* 185:587–602.

Bazin E, Glemin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animal. *Science* 312:570–572.

Beaumont M. 2010. Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Evol Syst.* 41:379–406.

Beaumont MA, Rannala B. 2004. The Bayesian revolution in genetics. *Nat Rev Genet.* 5:251–261.

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

Carnaval A, Hickerson M, Haddad CFB, Rodrigues MT, Moritz C. 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* 323:785–789.

Comes HP, Kadereit JW. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends Plant Sci.* 3:432–438.

Cook SR, Gelman A, Rubin DB. 2006. Validation of software for Bayesian models using posterior quantiles. *J Comput Graph Stat.* 15:675–692.

Cornuet J-M, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and micro-satellite data with the software DIYABC(v1.0). *BMC Bioinformatics* 11:401.

Csilléry K, Blum MGB, Gaggiotti OE, Francois O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol.* 25: 410–418.

Csilléry K, François O, Blum M. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol.* 3:475–479.

Cutler KB, Edwards RL, Taylor FW, Cheng H, Adkins J, Gallup CD, Cutler PM, Burr GS, Bloom AL. 2003. Rapid sea-level fall and deep-ocean temperature change since the last interglacial period. *Earth Planet Sci Lett.* 206:253–271.

Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* 292:673–679.

Dolman G, Joseph L. 2012. A species assemblage approach to comparative phylogeography of birds in southern Australia. *Ecol Evol.* 2: 354–369.

Donnellan SCA, Armstrong JA, Pickett MC, Milne TC, Baulderstone JC. 2009. Systematic and conservation implications of mitochondrial DNA diversity in emu-wrens, *Stipiturus* (Aves: Maluridae). *Emu* 109:143–152.

Drew JA, Barber PH. 2012. Comparative phylogeography in Fijian coral reef fishes: a multi-taxa approach towards marine reserve design. *PLoS One* 7:e47710.

Dynesius M, Jansson R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proc Natl Acad Sci U S A.* 97:9115–9120.

Edwards S. 1993. Mitochondrial gene genealogy and gene flow among island and mainland populations of a sedentary songbird, the grey-crowned babbler (*Pomatostomus temporalis*). *Evolution* 47:1118–1137.

Emerson B, Gillespie R. 2008. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol Evol.* 23: 619–630.

Emerson BC, Cicconardi F, Fanciulli PP, Shaw PJA. 2011. Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philos Trans R Soc B Biol Sci.* 366:2391–2402.

Excoffier L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol.* 13:853–864.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. Bayesian data analysis. Boca Raton (FL): Chapman & Hall/CRC.

Gillespie JH. 2001. Is the population size of a species relevant to its evolution. *Evolution* 55:2161–2169.

Graham RW, Lundelius EL Jr, Graham MA, Schroeder EK, Toomey RS III, Anderson E, Barnosky D, Burns JA, Churcher CS, Grayson DK, et al. 1996. Spatial response of mammals to late Quaternary environmental fluctuations. *Science* 272:1601–1606.

Grant WS, Bowen BW. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered.* 89:415–426.

Guay P-J, Chesser RT, Mulder RA, Afton AD, Paton DC, McCracken KG. 2010. East–west genetic differentiation in Musk Ducks (*Biziura lobata*) of Australia suggests late Pleistocene divergence at the Nullarbor Plain. *Conserv Genet.* 11:2105–2120.

Heller R, Chikhi L, Siegismund H. 2013. The confounding effect of population structure on bayesian skyline plot inferences of demographic history. *PLoS One* 8:e62992.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci.* 359:183.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD. 2010. Phylogeography's past, present, and future: 10 years after. *Mol Phylogenet Evol.* 54:291–301.

Hickerson MJ, Meyer C. 2008. Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evol Biol.* 8:322.

Hickerson MJ, Stahl EA, Lessios HA. 2006. Test for simultaneous divergence using approximate Bayesian computation. *Evolution* 60: 2435–2453.

Hickerson MJ, Stone GN, Lohse K, Demos TC, Xie X, Landerer C, Takebayashi N. 2014. Recommendations for using msBayes to incorporate uncertainty in selecting an abc model prior: a response to Oaks et al. *Evolution* 68:284–294.

Hill JK, Griffiths HM, Thomas CD. 2011. Climate change and evolutionary adaptations at species' range margins. *Annu Rev Entomol.* 56: 143–159.

HilleRisLambers J, Adler PB, Harpole WS, Levine JM, Mayfield MM. 2012. Rethinking community assembly through the lens of coexistence theory. *Annu Rev Ecol Evol Syst.* 43:227–248.

Huang W, Takebayashi N, Qi Y, Hickerson MJ. 2011. MTML-msBayes: approximate Bayesian comparative phylogeographic inference from multiple taxa and multiple loci with rate heterogeneity. *BMC Bioinformatics* 12:1.

Ho SYW, Gilbert MTP. 2010. Ancient mitogenomics. *Mitochondrion* 10: 1–11.

Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Resour.* 11: 423–434.

Ilves KL, Huang W, Wares JP, Hickerson MJ. 2010. Colonization and/or mitochondrial selective sweeps across the North Atlantic intertidal assemblage revealed by multi-taxa approximate Bayesian computation. *Mol Ecol.* 19:4505–4519.

Jeffreys H. 1961. Theory of probability. Oxford: Clarendon Press.

Joseph L, Adcock GJ, Linde C, Omland KE, Heinsohn R, Terry Chesser R, Roshier D. 2009. A tangled tale of two teal: population history of the grey *Anas gracilis* and chestnut teal *A. castanea* of Australia. *J Avian Biol.* 40:430–439.

Joseph L, Dolman G, Donnellan S, Saint KM, Berg ML, Bennett ATD. 2008. Where and when does a ring start and end? Testing the ring-species hypothesis in a species complex of Australian parrots. *Proc Biol Sci.* 275:2431–2440.

Joseph L, Wilke T. 2006. Molecular resolution of population history, systematics and historical biogeography of the Australian ringneck parrots *Barnardius*: are we there yet? *Emu* 106:49.

Joseph L, Wilke T, Alpers D. 2002. Reconciling genetic expectations from host specificity with historical population dynamics in an avian brood parasite, Horsfield's Bronze-Cuckoo *Chalcites basalis* of Australia. *Mol Ecol.* 11:829–837.

Joseph L, Zeriga T, Adcock G, Langmore N. 2011. Phylogeography and taxonomy of the Little Bronze-Cuckoo (*Chalcites minutillus*) in Australia. *Emu* 111:113–119.

Kass RE, Raftery AE. 1995. Bayes factors. *J Amer Stat Assoc.* 90:773–795.

Kearns AM, Joseph L, Cook LG. 2010. The impact of Pleistocene changes of climate and landscape on Australian birds: a test using the Pied Butcherbird (*Cracticus nigrogularis*). *Emu* 110:285.

Kearns AM, Joseph L, Edwards SV, Double MC. 2009. Inferring the phylogeography and evolutionary history of the splendid fairy-wren *Malurus splendens* from mitochondrial DNA and spectrophotometry. *J Avian Biol.* 40:7–17.

Kearns AM, Joseph L, Omland KE, Cook LG. 2011. Testing the effect of transient Plio-Pleistocene barriers in monsoonal Australo-Papua: did mangrove habitats maintain genetic connectivity in the Black Butcherbird? *Mol Ecol.* 20:5042–5059.

Knowles LL. 2009. Statistical phylogeography. *Annu Rev Ecol Evol Syst.* 40: 593–612.

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

Lavergne S, Mouquet N, Thuiller W, Ronce O. 2010. Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities. *Annu Rev Ecol Evol Syst.* 41: 321–350.

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.

Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK, Gilbert MTP, Nielsen R, et al. 2011. Species-specific responses of late Quaternary megafauna to climate and humans. *Nature* 479:359–364.

Marin J-M, Pillai N, Robert CP, Rousseau J. 2013. Relevant statistics for Bayesian model choice. *J Royal Stat Soc Series B.* Advance Access published December 9, 2013, doi: 10.1111/rssb.12056.

McLean AJ, Toon A, Schmidt DJ, Joseph L, Hughes JM. 2012. Speciation in chestnut-shouldered fairy-wrens (*Malurus* spp.) and rapid phenotypic divergence in variegated fairy-wrens (*Malurus lamberti*): a multilocus approach. *Mol Phylogenet Evol.* 63:668–678.

Mikheyev AS, Vo T, Mueller UG. 2008. Phylogeography of post-Pleistocene population expansion in a fungus-gardening ant and its microbial mutualists. *Mol Ecol.* 17:4480–4488.

Morgan K, O'Loughlin SM, Chen B, Linton Y-M, Thongwat D, Somboon P, Fong MY, Butlin R, Verity R, Prakash A, et al. 2011. Comparative phylogeography reveals a shared impact of pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Mol Ecol.* 20:4533–4549.

Murphy SA, Joseph L, Burbidge AH, Austin J. 2010. A cryptic and critically endangered species revealed by mitochondrial DNA analyses: the Western Ground Parrot. *Conserv Genet.* 12:595–600.

Nabholz B, Glémin S, Galtier N. 2008. Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Mol Biol Evol.* 25:120–130.

Navascués M, Emerson BC. 2009. Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Mol Ecol.* 18:4390–4397.

Nicholls JA, Austin JJ. 2005. Phylogeography of an east Australian wet-forest bird, the satin bowerbird (*Ptilonorhynchus violaceus*), derived from mtDNA, and its relationship to morphology. *Mol Ecol.* 14:1485–1496.

Pahnke K, Zahn R, Elderfield H, Schulz M. 2003. 340,000-year centennial-scale marine record of Southern Hemisphere climatic oscillation. *Science* 301:948–952.

Pavlova A, Amos JN, Joseph L, Loynes K, Austin JJ, Keogh JS, Stone GN, Nicholls JA, Sunnucks P. 2013. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution* 67:3412–3428.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol.* 16:1791–1798.

Prunier R, Holsinger KE, Carlson JE. 2012. The effect of historical legacy on adaptation: do closely related species respond to the environment in the same way? *J Evol Biol.* 25:1636–1649.

R Development Core Team. 2008. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Ramakrishnan U, Hadly EA. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Mol Ecol.* 18:1310–1330.

Ratnasingham S, Hebert P. 2007. BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes.* 7:355–364.

Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol.* 20:76–86.

Robert CP, Cornuet J-M, Marin J-M, Pillai NS. 2011. Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci.* 108:15112–15117.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol.* 9:552–569.

Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152:1079–1089.

Siddall M, Rohling E. 2008. Marine isotope stage 3 sea level fluctuations: data synthesis and new outlook. *Rev Geophys.* 46:1–29.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS. 2006. Comparative phylogeography of unglaciated eastern North America. *Mol Ecol.* 15:4261–4293.

Stone GN, Lohse K, Nicholls JA, Fuentes-Ultrilla P, Sinclair F, Schonrogge K, Csoka G, Melika G, Nieves-Aldrey J-L, Pujade-Villar J, et al. 2012. Reconstructing community assembly in time and space reveals enemy escape in a Western Palearctic insect community. *Curr Biol.* 22:532–537.

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

Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. 2013. Approximate Bayesian computation. *PLoS Comput Biol.* 9: e1002803.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 2002. Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol.* 7:453–464.

Tavare S, Balding D, Griffiths R, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.

Toon A, Austin JJ, Dolman G, Pedler L, Joseph L. 2012. Evolution of arid zone birds in Australia: leapfrog distribution patterns and mesic-arid connections in quail-thrush (*Cinclosoma*, Cinclosomatidae). *Mol Phylogenet Evol.* 62:286–295.

Toon A, Hughes JM, Joseph L. 2010. Multilocus analysis of honeyeaters (Aves: Meliphagidae) highlights spatio-temporal heterogeneity in the influence of biogeographic barriers in the Australian monsoonal zone. *Mol Ecol.* 19:2980–2994.

Weiher E, Freund D, Bunton T, Stefanski A, Lee T, Bentivenga S. 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philos Trans R Soc B Biol Sci.* 366: 2403–2413.

Wood DA, Vandergast AG, Barr KR, Inman RD, Esque TC, Nussear KE, Fisher RN, Bode M. 2012. Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Divers Distrib.* 19:722–737.

Yu Z, Eicher U. 1998. Abrupt climate oscillations during the last deglaciation in central North America. *Science* 282:2235–2238.