MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 5059–5072 doi: 10.1111/j.1365-294X.2012.05747.x

Testing hypotheses for genealogical discordance in a rainforest lizard

SONAL SINGHAL*†[1] and CRAIG MORITZ*†‡

*Museum of Vertebrate Zoology, University of California, 3101 Valley Life Sciences Building, Berkeley, CA, 94720-3160, USA, †Department of Integrative Biology, University of California, 1009 Valley Life Sciences Building, Berkeley, CA, 94720-3140, USA and ‡Division of Evolution, Ecology & Genetics, Research School of Biology, The Australian National University, Canberra, ACT OZOO, Australia

Abstract

Genealogical discordance, or when different genes tell distinct stories although they evolved under a shared history, often emerges from either coalescent stochasticity or introgression. In this study, we present a strong case of mito-nuclear genealogical discordance in the Australian rainforest lizard species complex of *Saproscincus basiliscus* and *S. lewisi*. One of the lineages that comprises this complex, the Southern *S. basiliscus* lineage, is deeply divergent at the mitochondrial genome but shows markedly less divergence at the nuclear genome. By placing our results in a comparative context and reconstructing the lineages' demography via multilocus and coalescent-based approximate Bayesian computation methods, we test hypotheses for how coalescent variance and introgression contribute to this pattern. These analyses suggest that the observed genealogical discordance likely results from introgression. Further, to generate such strong discordance, introgression probably acted in concert with other factors promoting asymmetric gene flow between the mitochondrial and nuclear genomes, such as selection or sex-biased dispersal. This study offers a framework for testing sources of genealogical discordance and suggests that historical introgression can be an important force shaping the genetic diversity of species and their populations.

Keywords: approximate Bayesian computation, cytonuclear discordance, demographic reconstruction, introgression, phylogeography, *Saproscincus basiliscus*

Received 2 December 2011; revision received 27 June 2012; accepted 1 July 2012

Introduction

Genealogical discordance is a common phenomenon in natural systems (McGuire *et al.* 2007; Melo-Ferreira *et al.* 2011; Near *et al.* 2011; Reid *et al.* 2012), yet the causes and consequences of discordance are often unclear. Under genealogical discordance, not all loci appear to tell the same story, even though the genes evolved under a common demographic history. This discordance can take several forms; most notably, topologies and branch lengths among organismal lineages can vary across loci (Jennings & Edwards 2005; Edwards 2009). Of note is discordance between organelle (chloroplastic and mitochondrial genomes) and nuclear loci, which appears throughout the natural world (Chan & Levin 2005; Petit & Excoffier 2009). Studies often point to the special characteristics of the organelle genome—i.e. its smaller effective population size, uniparental inheritance, lack of recombination, key role in organismal metabolism and, in the case of the mitochondrial genome, increased mutation rate (Ballard & Whitlock 2004)—to explain this discordance. However, it is unclear if the special characteristics of cytoplasmic genomes need to be invoked to explain cytonuclear discordance as compared to general genealogical discordance (Currat *et al.* 2008).

Whether in the form of discrepancies in topology or branch lengths, genealogical discordance typically arises from three, nonexclusive processes: coalescent variance ('incomplete lineage sorting'), introgression and gene

[1]Correspondence: Sonal Singhal, Fax: 510 643 8238; E-mail: singhal@berkeley.edu

duplication (Maddison 1997). Here, we focus on coalescent variance and introgression. First, the coalescent, or the process by which alleles in a population find a common ancestor, is inherently stochastic (Wakeley 2008). Thus, theory predicts that any genealogical reconstruction should exhibit some heterogeneity across loci—not only because the coalescent is a sampling process but also because we rely on the distribution of mutations—another stochastic process, to estimate coalescent histories (Wakeley 2008). How much heterogeneity is expected is unclear; for a subset of population histories, researchers have derived analytical expectations for the variance in coalescent times and inferred genealogical relationships (Tavaré *et al.* 1997; Slatkin & Pollack 2008). However, this work is generally limited to simple splitting histories, and more complex patterns of divergence could possibly increase this variance (Lenormand *et al.* 2009). A second powerful source of genealogical discordance is introgression, or the movement of an allele from one gene pool to another (Anderson 1949; Kuo & Avise 2005). Particularly when introgression acts in concert with other forces such as locus-specific selection or sex-biased dispersal, genealogical discordance can increase further (Maroja *et al.* 2009).

Both coalescent variance and introgression are often invoked by researchers trying to explain patterns of genealogical discordance (Near *et al.* 2011). Although not always easy (Melo-Ferreira *et al.* 2012; Reid *et al.* 2012), it is important to determine how these forces interact in the context of a species' history to create genealogical discordance. After all, discordance is useful for inferring key parameters about the divergence process (Edwards & Beerli 2000); in particular, discordance increases with larger ancestral population sizes (Hey 2010). Further, because it can result from introgression, discordance can inform us about both historical and ongoing hybridization in the lineages of interest (Near *et al.* 2011). Thus, although genealogical discordance has sometimes been regarded as a complication (Degnan & Rosenberg 2009), it actually can provide an informative window into species' histories.

In this study, we present a compelling case of genealogical discordance in the rainforest lizard species complex, *Saproscincus basiliscus* and *S. lewisi*. Endemic to the Australian Wet Tropics (AWT), a narrow strip of rainforest in north-eastern Australia, this complex consists of four major, highly divergent mitochondrial lineages: *S. lewisi* in the far north, and the Northern, Central and Southern lineages of *S. basiliscus* (Fig. 1A, B; Moussalli *et al.* 2009). Throughout most of their history, paleomodelling suggests that the lineages likely diverged in isolated glacial-period refugia, with brief opportunities for gene flow during interglacial periods (Moussalli *et al.* 2009; VanDerWal *et al.* 2009). Here, to further explore the divergence history of these regional populations, we collect a species-wide and multilocus data set. In doing so, we find significant branch length heterogeneity between the nuclear and mitochondrial genomes for the Southern-most populations—50-fold greater divergence of mitochondrial DNA than nuclear DNA. To test hypotheses about how coalescent variance and introgression contribute to this discordance, we place the data in a comparative context, and we exploit the region's well-understood biogeography to reconstruct the lineages' demographic history via approximate Bayesian computation (ABC) analyses.

Methods

Sampling and genetic data

Our sampling covers the known distribution of the sister species *Saproscincus basiliscus* and *S. lewisi* throughout the AWT (Moussalli *et al.* 2009). These species are leaf-litter skinks that, while generally found in association with rainforest, can also extend into adjacent wet sclerophyll forests (Couper & Keim 1998). The two species are ecologically and morphologically similar; they are delineated based on a minor morphological feature—differing number of paravertebral scales (Couper & Keim 1998).

In this study, we added to the mitochondrial data set collected by Moussalli *et al.* (2009) to generate a 291-individual data set. Our expanded sampling focused on sequencing individuals located in geographic gaps between previously defined phylogeographic lineages; habitat is contiguous between all lineages, except for the Central and Southern lineages (Fig. 1A). For a subset of these individuals ($n = 86$; Table S1), we sequenced eight nuclear loci. Because the relevant unit of analysis in this study is the phylogeographic lineage, these individuals were sampled roughly proportional to the prevalence of each mitochondrial clade, and we ensured representation of the full geographic range of the species.

We extracted DNA from preserved tail tissue using a high-salt DNA extraction (Aljanabi & Martinez 1997). To assay mitochondrial variation, we sequenced the ND4 locus (Arevalo *et al.* 1994); to assay nuclear variation, we sequenced eight loci, including six published previously: β-globin intron, C-mos exon, R35 exon, Rhodopsin intron, and TPI and RPS8 introns (Saint *et al.* 1998; Dolman & Phillips 2004; Leaché 2009; Bell *et al.* 2010). To this, we added two additional intronic loci (CRISP and LGMN), designed for the closely related lizard *Lampropholis coggeri* (Bell *et al.* 2010). All PCRs were conducted in standard conditions in 12 μL volumes, using a touchdown protocol of 14 cycles of decremental and 22 cycles of stable annealing temperatures (details available in Table S2). Following PCR amplification, we visualized products on an agarose gel, cleaned PCR

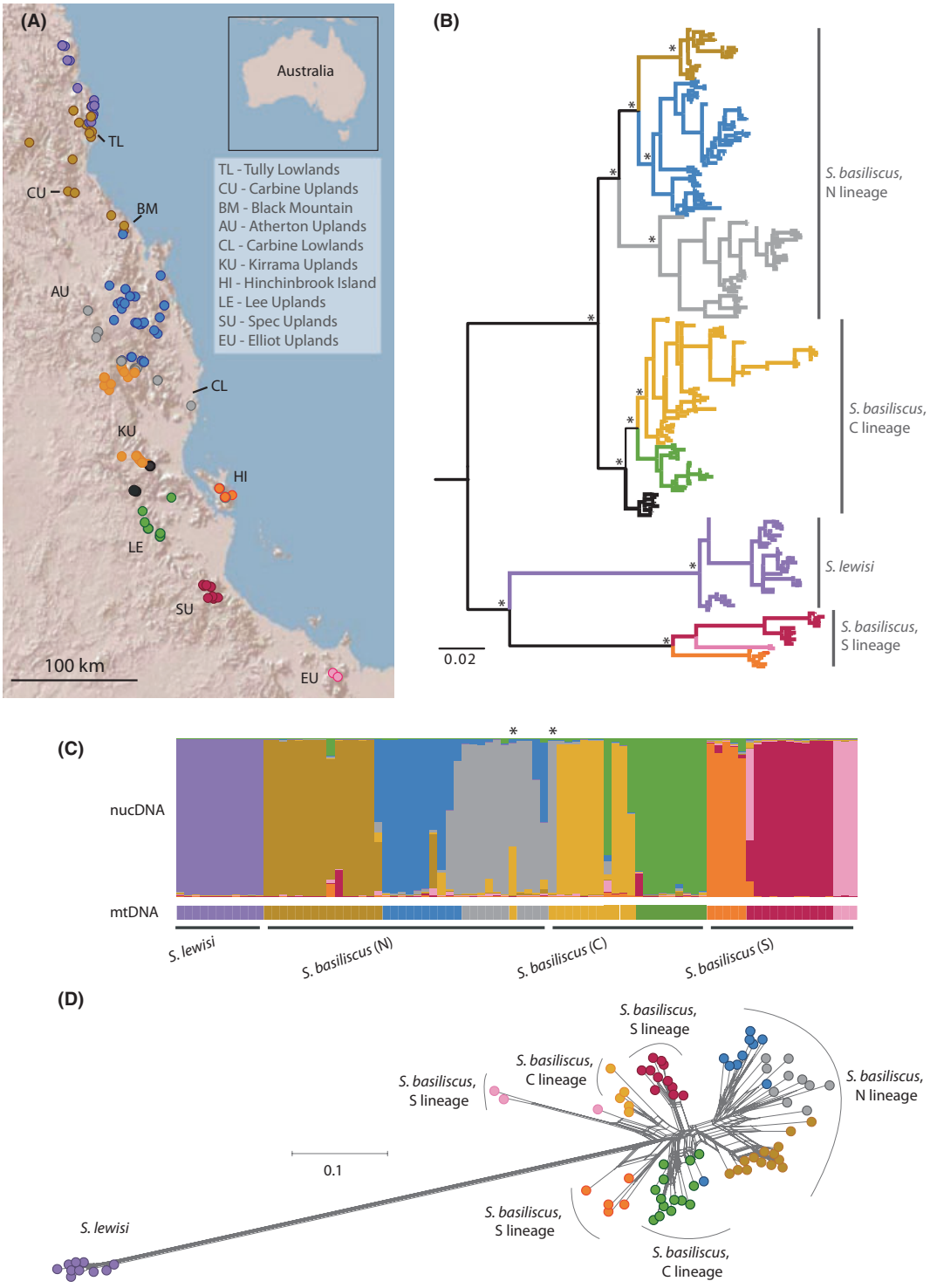

Fig. 1 (A) Map of Australian Wet Tropics showing sampled points for *Saproscincus basiliscus* and *S. lewisi* and identifying bioregions (Williams *et al.* 1996). (B) Mitochondrial gene tree as inferred by Bayesian analysis for *S. basiliscus* and *S. lewisi*. Major clades with posterior probability >0.95 are marked with an asterisk. (C) Structure results based on haplotypes at eight nuclear loci, with mitochondrial identity for each individual shown. Individuals are ordered by location from north to south. Instances of mito-nuclear discordance are identified by asterisks. (D) Gene network based on haplotypes at eight nuclear loci. Scale is in a standardized, nonunit-based measurement given by POFAD.

products via ExoSAP-IT (USB) and sequenced products using BigDye v3.1 on an ABI3730 (Applied Biosystems). The majority of reads were assembled and edited using Geneious (Drummond *et al.* 2010); to resolve assemblies with heterozygous indels, we used CodonCode Aligner's heterozygoteIndel feature (CodonCode Co.).

For the mitochondrial locus, final assemblies were aligned with the published alignment from Moussalli *et al.* (2009) using MUSCLE (Edgar 2004). For the nuclear loci, we inferred haplotypes from our diplotypes computationally using PHASE2.1 (Stephens *et al.* 2001), running the algorithm 100 times and assuming a constant recombination rate. We used the most probable haplotype resolution to determine haplotypes; here, all heterozygous sites were resolved to >95% probability. Final nuclear alignments were made with MUSCLE and checked manually in Geneious. The final nuclear data set was 94% complete by locus with a combined length of 3.98 kb.

Tree-based and multilocus analyses

To infer genealogical relationships among our mitochondrial haplotypes, we used the Bayesian phylogenetic approach implemented in MrBayes v3.1.2 (Huelsenbeck & Ronquist 2001). We included sequences from the species *Saproscincus czechurai* and *S. tetradactylus* as outgroups (Moussalli *et al.* 2009; GenBank IDs: FJ195325.1, J195291.1). We partitioned the alignment into the genic and tRNA regions and assigned to each partition the most appropriate model for nucleotide substitution using MrModelTest (Nylander 2004). The partitioned alignment was run twice, each with four chains (three heated, one cold, default heating parameters), for 20 000 000 generations with a 6 000 000 generation burn-in. MCMC chain convergence was assessed by calculating ESS values using Tracer, and the posterior distribution of trees was summarized using TreeAnnotator (Drummond & Rambaut 2007).

To describe heterogeneity in topologies and branch lengths among loci, we inferred gene trees for each of our nuclear loci using the maximum-likelihood approach implemented in RAxML (Stamatakis 2006). Each alignment was run unpartitioned, under the substitution model inferred to be most probable by MrModelTest (Nylander 2004). If the model selected by MrModelTest was simpler than those implemented in RAxML, we chose the simplest model RAxML provides (GTRGAMMA). For each locus, we found the best-scoring maximum-likelihood tree and conducted 1000 rapid bootstrap analyses to determine support for the tree.

We employed two approaches to summarize and visualize patterns of variation across our multilocus results. First, to identify population clusters, we used the program Structure v2.3.2, which identifies populations (K) by minimizing linkage and Hardy–Weinberg disequilibrium within a cluster (Pritchard *et al.* 2000). We ran Structure 20 times with our phased nuclear data (10 000 000 steps with 1 000 000 burn-in) under the 'admixture' model for each of 12 K values (ranging from 1 to 12), determined the best-supported K value following Evanno *et al.* (2005) as implemented in StructureHarvester (Earl 2011), and summarized and plotted results using Clumpp and Distruct (Rosenberg 2004; Jakobsson & Rosenberg 2007). Second, we used POFAD to construct a network of individual similarity based on phased nuclear data. POFAD is a distance-based method that explicitly accounts for haplotypic variation within individuals (Joly & Bruneau 2006). We inferred Tamura-Nei corrected distance matrices for each locus using PAUP (Swofford 2002), calculated a final individual-based distance matrix with POFAD and visualized the results as unrooted networks using SplitsTree (Hudson & Bryant 2006).

Between-lineage diversity

Discrepant branch lengths for mitochondrial and nuclear loci could simply result from differences in mutation rates between the two genomes. If mutational variance is a minor factor in this system, we would expect sequence divergence at mitochondrial and nuclear genomes to be correlated and to reflect the difference in substitution rate between the genomes. To explore this possibility, we used Arlequin v3.1 to estimate raw D_{xy} and net D_a sequence divergence between the lineages of the *S. basiliscus* species complex (Fig. 1B) at both the mitochondrial and nuclear genomes, as estimated by the Tamura–Nei model. To compare patterns of nuclear-mitochondrial divergence more generally, we expanded our analysis of divergence between lineages to five other closely related and co-distributed species of lizards: *Carlia rubrigularis* and *C. rhomboidalis* (7 nuclear loci; Dolman & Moritz 2006), *L. robertsi* (8 nuclear loci; Bell *et al.* 2010), *L. coggeri* (6 nuclear loci; Bell *et al.* 2010), and *Gnypetoscincus queenslandiae* (2 nuclear loci; Singhal, unpublished). Many of the same nuclear loci were sequenced for these species and this study's species, and the same mitochondrial marker was sequenced across all species. In each of these species complexes, we identified major lineages based on the mitochondrial genealogy and then determined sequence divergence between sister phylogeographic lineages at both mitochondrial and phased nuclear data.

Divergence and demographic analyses

To determine if coalescent variance or introgression could explain the observed, strong genealogical discordance, we inferred the most likely demographic history

using ABC. For many demographic scenarios, both defining and calculating the likelihood function for the model can be challenging; the crude approximation afforded by ABC, can estimate the likelihood function by simulation and, thus, can allow researchers to test a much wider range of biologically relevant models (Beaumont 2010). That being said, recent research has suggested that model choice via ABC can give slightly biased results for un-nested models, particularly when summary statistics fail to capture the full complexity of the raw data (Robert *et al.* 2011). However, this and subsequent research also indicate, by using summary statistics with differing distributions under alternative models and by validating the model choice procedure itself (Marin *et al.* 2011), ABC can remain a powerful tool for exploring and testing the fit of different models to data.

With these caveats in mind, we tested the fit of two major classes of models to our empirical data: models without introgression, designed to test how coalescent variance can contribute to generating genealogical discordance, and those with introgression, designed to test how introgression, together with coalescent variance, can contribute to generating genealogical discordance. Our choice of models is motivated by our knowledge of biogeographic history of the *S. basiliscus* species complex and its rainforest habitat (Moussalli *et al.* 2009; VanDerWal *et al.* 2009). We focus on modelling the sister lineages that exhibit strong genealogical discordance, the Central and Southern lineages of *S. basiliscus*. To simplify the models, we model only the Spec Uplands population of the Southern lineage; the Hinchinbrook Island and Elliot Upland populations have too low of sample sizes to model accurately and also introduce additional spatial complexity beyond the scope of this study. As predicted by paleomodels (Fig. 2A), the Central and Southern lineages likely evolved largely in allopatry through glacial cycles, and thus we consider multiple variations on a basic allopatric model in our simulations (Fig. 3). These models are:

- Models with no introgression

1 A simple model of population splitting, in which an ancestral population splits into the Central and Southern lineages with no postdivergence gene flow (Fig. 3A).
2 An extension of the simple model ('peripatric divergence') in which the Southern lineage is initially very small when it splits from the ancestral lineage (Fig. 3B).
3 A model in which there is ancestral population structure, such that the ancestral population consists of multiple populations with limited gene flow, after which it splits into the Central and Southern lineages (Fig. 3C).

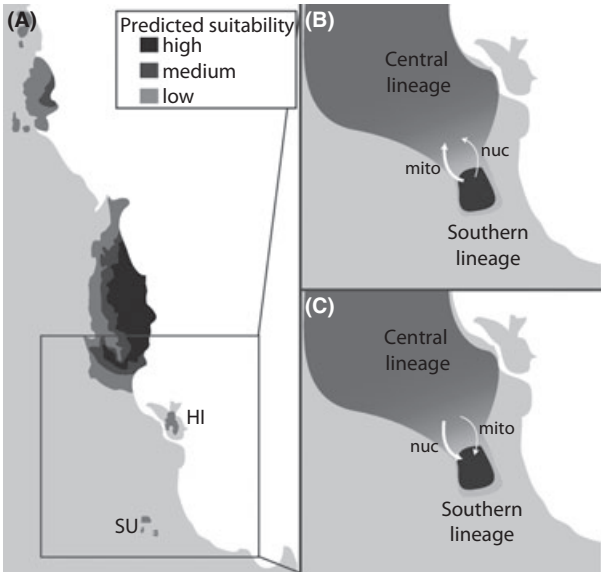

Fig. 2 (A) A suitability map for wet sclerophyll rainforest in the Australian Wet Tropics showing isolated glacial refugia during the cold–dry stage of the glacial cycle (18 000 ybp), modified from VanDerWal *et al.* (2009). Cartoon depictions of the possible historical introgression event during a cool-wet period of higher connectivity, showing the dynamics of mitochondrial and nuclear introgression between the Central and Southern lineages of *Saproscincus basiliscus* for (B) model 5, of more mitochondrial than nuclear gene flow and (C) model 6, of more nuclear than mitochondrial gene flow.

- Models with introgression

4 A model in which there is a pulse of postdivergence gene flow, in which the Central lineage expands and exchanges migrants with the Southern lineage for a brief period of time in the past (Fig. 3D).
5 A model in which there is a pulse of postdivergence gene flow, in which the mitochondrial gene flow between the Central and Southern lineages is greater than nuclear gene flow (Fig. 3E). This model would allow the relictual Southern mitochondrial genome to introgress into the invading Central lineage (Fig. 2B).
6 A model in which there is a pulse of postdivergence gene flow, in which nuclear gene flow between the Central and Southern lineages is greater than mitochondrial gene flow (Fig. 3F). This model would allow the invading Central nuclear genome to completely introgress the relictual Southern population (Fig. 2C).

In particular, we include models 2 and 3 because such histories can substantially affect the coalescent (Jesus *et al.* 2006; Slatkin & Pollack 2008), and they are

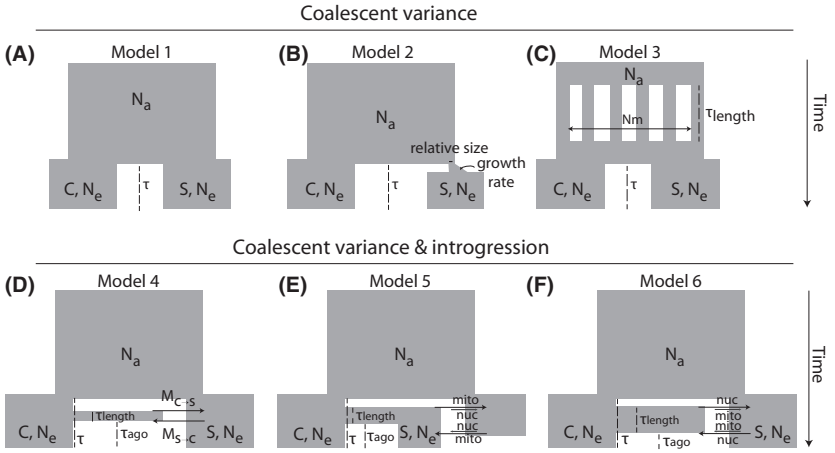

Fig. 3 Cartoon depictions of the six models used in the approximate Bayesian computation (ABC) analysis to infer the divergence history of the Central and Southern lineages of *Saproscincus basiliscus*: (A) simple splitting, (B) a 'peripatric' splitting model, (C) a splitting model with ancestral population structure, (D) a model with pulsed, postdivergence gene flow, (E) a model with pulsed gene flow in which mitochondrial gene flow is greater than nuclear gene flow, and (F) a model with pulsed gene flow in which nuclear gene flow is greater than mitochondrial gene flow.

plausible in context of the biogeographic history of these lineages (Moritz *et al.* 2009). The introgression models (models 4–6) reflect our knowledge of this system's history; a model in which there is constant gene flow throughout divergence is unlikely to describe this system. Reconstruction of the AWT rainforest during the Pleistocene suggests that the Central and Southern lineages were isolated for most of their history—including at present (Moussalli *et al.* 2009)—with brief periods of increased connectivity during cool-wet periods, most recently in a brief period during the early Holocene (Fig. 2; VanDerWal *et al.* 2009). Further, as has been suggested in numerous other systems of cytonuclear discordance (Ballard & Whitlock 2004; Good *et al.* 2008), models 5 and 6 explore the possibility of differential rates of gene flow at the mitochondrial and nuclear genomes if, for example, there is allele surfing (Currat *et al.* 2008), the mitochondrial genome is under selection (i.e. cytonuclear incompatibilities), or there is sex-biased gene flow (Discussion). Importantly, model 5 differs from models 4 and 6 in the structure of the simulation (Fig. 3D–F), because, although the geography of all three scenarios is the same, the relative movement of mitochondrial and nuclear genomes between the Central and Southern lineages differs across models (Fig. 2B, C).

We used the program MSABC to simulate and generate summary statistics for each of these models, modifying the program via Perl scripts to both simulate nuclear and mitochondrial data and to calculate six additional summary statistics ($D_{a,\mathrm{nuc}}$, $D_{a,\mathrm{mito}}$, $D_{xy,\mathrm{nuc}}$, $D_{xy,\mathrm{mito}}$) and the corresponding cytonuclear divergence ratios ($\frac{D_{a,\mathrm{mito}}}{D_{a,\mathrm{nuc}}}$, $\frac{D_{xy,\mathrm{mito}}}{D_{xy,\mathrm{nuc}}}$). For loci lengths, mutation rate and recombination rate, we used well-circumscribed priors defined by our empirical data and data from other studies of lizards (Rosenblum *et al.* 2004; Brandley *et al.* 2011); for all other parameters, we used broad, uninformative priors (Table S3).

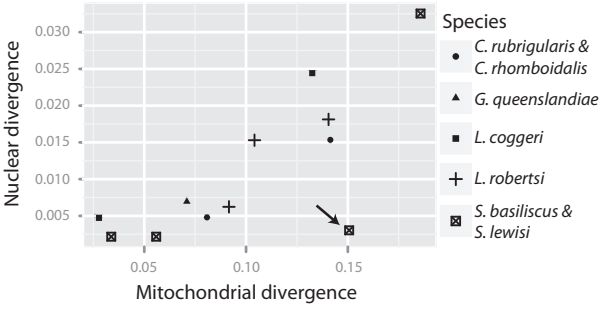

Fig. 4 Correlation between mitochondrial and nuclear divergence between phylogeographic lineages in Australian Wet Tropics lizards; points are labelled according to the species in which the contact is found. The arrow identifies the contact of interest: *Saproscincus basiliscus* Central/Southern lineages.

We generated an initial set of 10 000 simulations under each model and used the results from these simulations to evaluate which summary statistics differed the most between our two major model classes (c.f. Robert *et al.* 2011) and to determine which statistics, if any, were significantly correlated and, thus, unlikely to provide additional information. Through this approach, we defined three summary statistics ($D_{a,\mathrm{nuc}}$, $D_{a,\mathrm{mt}}$, $\frac{D_{a,\mathrm{mito}}}{D_{a,\mathrm{nuc}}}$), and we used these along with summary statistics more generally useful for inferring demography ($\theta_{w,\mathrm{nuc}}$ and $\theta_{w,\mathrm{mito}}$ for both lineages). We then simulated larger data sets of 1 million simulations for each model to use

in model choice. Using the R package ABC (Csilléry *et al.* 2012), we conducted all downstream inference. Primarily, we used a weighted multinomial logistic regression to estimate the posterior probabilities of our models' fit to our data. Following a rejection step across all models (tolerance rate; $\gamma = 0.01$), regression was performed on the retained simulations, where the model is treated as a categorical response variable and the summary statistics are the independent variables (Beaumont 2008). Additionally, while our primary objective here is model choice rather than model fitting, we used model fitting to test the accuracy of model choice (Gelman *et al.* 2004). Thus, we then inferred the posterior distributions for each parameter in each demographic model using a local linear regression-corrected rejection scheme ($\gamma = 0.01$) and log-transforming parameters prior to fitting to ensure the posterior distributions fell within prior ranges (Hamilton *et al.* 2005).

Finally, we evaluated the performance of our model choice procedure via two methods. First, we generated pseudo-observed data sets, for which we randomly selected a simulated data set, defined which model supported the data best using the same model choice procedure described above and then calculated how often these data sets were mis-classified. Second, we generated posterior predictive distributions under the inferred posterior distributions for each model and then compared our empirical summary statistics to these simulated distributions; if model choice is accurate, our empirical summary statistics should lie within the simulated distributions (Thornton & Andolfatto 2006).

Results

Mitochondrial and multilocus phylogeography

The mitochondrial gene tree recovered the same lineages as described by Moussalli *et al.* (2009; Fig. 1B). Here, we refer to these mitochondrial lineages as the *S. lewisi*, Northern, Central and Southern lineages. Each major mitochondrial lineage consists of several, well-supported subclades, each of which is geographically restricted (Fig. 1B). Our improved sampling located areas of sympatry between the *S. lewisi* and Northern lineages and between the Northern and Central lineages. As described earlier, *S. lewisi* and Southern *S. basiliscus* are each highly divergent from the rest of the clade; net corrected sequence divergence ranges from 15% to 18% between these lineages. Further, the Southern lineage is highly structured; populations in the currently isolated Spec Uplands, Elliot Uplands and Hinchinbrook Islands are 3–8% divergent from each other (Fig. 1A, B).

The primary result found from analyses of the eight-loci nuclear data is marked genealogical discordance between the mitochondrial and nuclear data for the *S. basiliscus* Southern lineage. As shown by the multilocus nuclear gene network generated via POFAD (Fig. 1D), *S. lewisi* is quite divergent from the rest of the species complex ($D_a = 0.032$). However, the *S. basiliscus* Southern lineage shows an order of magnitude less nuclear divergence from the Central and Northern lineages ($D_a = 0.003$), even though its mitochondrial divergence is nearly as great as that of *S. lewisi*.

Besides this instance of genealogical discordance, the data otherwise show broadscale concordance among markers. Although individual gene trees all show differences in topology and branch length for most of these lineages (Fig. S1), a consensual history emerges from multilocus analyses. First, the major clades and subclades identified by mitochondrial sequencing are all recovered by Structure clustering of nuclear genotypes (Fig. 1C). Using the Evanno method (Evanno *et al.* 2005), we determined that nine clusters provided the best fit to the data, each of which correspond to a mitochondrial clade. Examination of the nuclear data shows, however, that only the *S. lewisi* lineage is recovered as a distinct clade, even though the main mitochondrial lineages are largely separated in the distance-based network (Fig. 1D). Second, geographic concordance between the two marker types is strong; there are only two cases where an individual belongs to two different major clusters (*S. basilicus* N, C, S lineages, and *S. lewisi*) for mitochondrial and nuclear data (Fig. 1C). Both cases trace to individuals sampled at the parapatric boundaries between the Northern and Central lineages.

Between-lineage diversity

To place our finding of incongruent branch lengths between the nuclear and mitochondrial genomes for the Southern lineage in context, we compared sequence divergence (D_a) in mitochondrial and nuclear genomes between major phylogeographic lineages in seven co-distributed species of lizards. As shown in Fig. 4, divergence levels at the two genomes are highly correlated ($r^2 = 0.91$; $P < 0.005$), with an average divergence ratio of 11.2. As expected, this divergence ratio reflects the estimate of the nuclear-mitochondrial substitution scalar for lizards (approximately 14; Brandley *et al.* 2011). The mito-nuclear divergence ratio between the *S. basiliscus* Southern and Central lineages (50.4; as identified by the arrow in Fig. 4) is a noticeable outlier in this group. Including this datum substantially weakens the strength of the correlation between nuclear and mitochondrial divergence ($r^2 = 0.76$; $P < 0.05$). Thus, the discordance we see in branch lengths between the

nuclear and mitochondrial genomes for the Southern lineage is unlikely due to variance in mutation rates.

Divergence and demographic analyses

To determine if coalescent stochasticity or introgression better explains the genealogical discordance we see, we used an ABC approach to test the fit of six different models to our empirical data. First, we used a small number of simulations to test the utility of a wide range of summary statistics to distinguish between our models. Perhaps because our models explore a large parameter space—portions of which lead to competing models becoming nearly identical—many of the tested summary statistics showed little difference in distributions between the models. That said, we identified seven summary statistics [($D_{a,\mathrm{nuc}}$, $D_{a,\mathrm{mt}}$, $\frac{D_{a,\mathrm{mito}}}{D_{a,\mathrm{nuc}}}$), $\theta_{w,\mathrm{nuc}}$ and $\theta_{w,\mathrm{mito}}$ for both lineages], which were not strongly correlated with each other ($r^2 < 0.2$) and that showed differing distributions between the models (Fig. S2).

We extended these initial simulations to conduct model choice by calculating posterior probabilities of our differing models. The general class of models exploring coalescent stochasticity (models 1 through 3) were supported with low posterior probability (summed $P = 0.0483$), and models invoking a pulse of gene flow were strongly supported (summed $P = 0.951$; Table 1; Bayes Factor = 19.68). In particular, within models with introgression, a model that allowed for asymmetric gene flow between the two genomes (summed $P = 0.922$) was very strongly supported compared with a model with equal gene flow between the two genomes ($P = 0.0291$; Bayes Factor = 31.68). Further, there is some support that a model with more mitochondrial gene flow (model 5; $P = 0.763$) is more likely than a model with more nuclear gene flow (model 6; $P = 0.159$; Bayes Factor = 4.79). Given the results of Robert *et al.* (2011), we refrain from over-interpreting the posterior probabilities reported here, but we do suggest that these results strongly support a demographic history of pulsed gene flow that is heterogenous between the nuclear and mitochondrial genomes.

Table 1 Models, numbered as in text, with posterior probabilities as inferred from approximate Bayesian computation analyses

Model	Posterior probability
Simple split (model 1)	0.0263
Peripatric split (model 2)	0.0220
Ancestral structure split (model 3)	0.000
Pulsed gene flow (model 4)	0.0291
More mitochondrial gene flow (model 5)	0.763
More nuclear gene flow (model 6)	0.159

We estimated the posterior distributions of the parameters for our best-fitting model (model 5; pulsed gene flow with greater mitochondrial than nuclear gene flow). Our results, shown in Fig. S3, are plausible given our knowledge of the species' ecology and the region's biogeographic history. Model fitting showed very little power to estimate key parameters of the pulse of gene flow, that is, when it started, how long it lasted, number of migrants in either direction or the magnitude of asymmetry in gene flow between the two genomes. Other studies looking at pulsed gene flow have shown similarly limited power (Li *et al.* 2010; Yeung *et al.* 2011); fortunately, our conclusions do not depend on precise estimates of these parameters.

To evaluate the performance of our model choice, we generated pseudo-observed data sets and looked at the frequency of mis-classification. As seen in Fig. S4 and Table S4, the frequency of false positives and negatives is high across most of the models. However, for most of these mis-classified models, the posterior probability of the best-supported model was low ($P < 0.5$). Following Fagundes *et al.* 2007, we computed the probability that the best-supported model is the correct model, given the observed posterior probability. Comparing models with and without introgression, we computed the probability of observing our posterior probability ($P = 0.951$) in error as 0. Comparing models with and without heterogeneous introgression, we computed the probability of observing our posterior probability ($P = 0.922$) in error as 0. Comparing a model with more mitochondrial gene flow than nuclear gene flow, we computed the probability of observing our posterior probability ($P = 0.763$) in error as 0.122. We further evaluated our model choice procedure by comparing the empirical value for our summary statistics to the posterior predictive distributions inferred for each model. For our best-supported model (model 5), each of our observed values is within the posterior predictive distributions (Fig. 5); the probability of recovering the same or more extreme value for a given summary statistic ranged from 0.129 to 0.724. For the other five models, this is only also true for model 6 (probabilities ranged from 0.074 to 0.823; Fig. S5). These evaluations of our model choice procedure strongly support model 5 or model 6 as the best fit to our data and support the hypothesis that historical introgression, with asymmetry in gene flow between the mitochondrial and nuclear genomes, is likely the source of the discordance seen in *S. basiliscus* and *S. lewisi*.

Discussion

By collating a multilocus data set in the species complex *Saproscincus basiliscus* and *S. lewisi*, we uncovered a

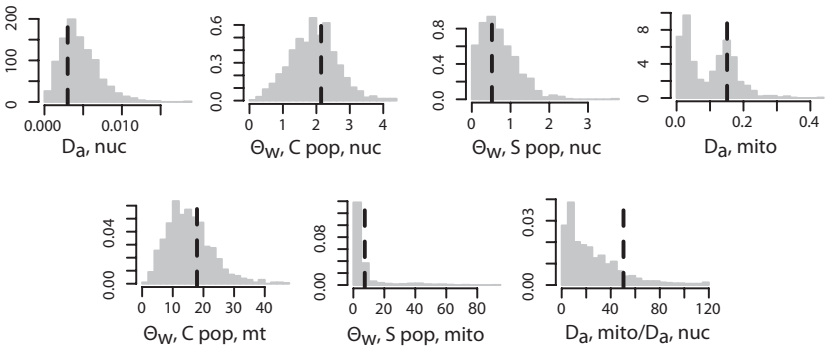

Fig. 5 Following approximate Bayesian computation (ABC) analysis of *Saproscincus basiliscus* Central and Southern lineages, posterior predictive results for the most highly supported model (model 5: greater pulsed gene flow at the mitochondrial genome) across all seven summary statistics. Dashed black lines reflect true value of summary statistic for the empirical data.

striking example of genealogical discordance: the populations representing *S. basiliscus* Southern mitochondrial lineage, which are 15% divergent from the rest of *S. basiliscus*, are an order of magnitude less divergent at the nuclear genome than expected. That *S. lewisi*, a lineage with similar levels of mitochondrial divergence, exhibits genealogical concordance across loci (Fig. 1; Fig. S1) and is reproductively isolated from the rest of the clade based on multilocus data from sympatric populations (Singhal and Moritz, unpublished), further underlines how anomalous this result is.

Perhaps the most parsimonious explanation for reduced divergence in the nuclear genome when compared with the mitochondrial genome is that the nuclear genome has a lower substitution rate. Calculating divergence between phylogeographic lineages in seven species of closely related AWT skinks shows that divergence at the mitochondrial genome is tightly correlated with nuclear divergence (Fig. 4). Although the substitution rates between the two genomes are different—the mito-nuclear divergence rate is approximately 11:1—our discordant clade has a divergence ratio of 50, far beyond what could be explained by genome-specific substitution profiles.

Source of genealogical discordance

What, then, can explain this level of genealogical discordance? Genealogical discordance is most often attributed to either the stochasticity of the coalescent or introgression across lineage boundaries (Kingman 1982). With respect to coalescent variance, because the mitochondrial genome acts as a single locus (Ballard & Whitlock 2004), it might be capturing an errant view of history, relative to the rest of the genome. If coalescent variance is underpinning genealogical discordance, we might expect that the population diverged under a history that promoted increased variance in genealogical patterns. Our ABC simulations, which explored both a broad set of divergence histories and parameter space, suggest under certain population histories, genealogical discordance (as measured here) can increase. However, the increase in variance is limited compared with the magnitude of the discrepancy we see (Fig. S2). Thus, as supported by the low posterior probabilities for these models, coalescent variance is unlikely the source of this system's genealogical discordance.

Our divergence history reconstruction suggests the genealogical discordance likely results from historical introgression between the Central and Southern lineages. Ongoing introgression between the Central and Southern lineages is unlikely; niche models and field surveys suggest the two lineages do not currently meet (Moussalli *et al.* 2009), and nuclear data show no evidence for recent admixture between these lineages (Fig. 1C). However, niche models for the preferred habitat of the species complex (e.g. wet sclerophyll/rainforest) through time support a model of divergence in isolation during glacial periods with transient connectivity and, thus, opportunities for introgression in the early Holocene (VanDerWal *et al.* 2009). Habitat during the cold–dry portion of the glacial cycle was predicted to be restricted to two major refugia in the north and south of the AWT (corresponding to the Northern and Central lineages, respectively), with smaller refugia in the Spec Uplands and on Hinchinbrook Island and limited interconnectivity between the refugia (Fig. 2A). During the cool-wet and warm-wet portions of the glacial cycles, the forest between the Central and Southern lineages was predicted to be contiguous. As such, the Central lineage likely expanded out of its refugium and invaded the southern AWT, allowing for gene flow between the Central and Southern lineages (Fig. 2B, C). Following gene flow between the two lineages, the ancestral Southern nuclear genome was replaced by the invading Central nuclear genome, but the highly

divergent Southern mitochondrial genome persisted. In a sense, the only evidence for the Southern population that persisted through time in the small southern refugia is its highly divergent mitochondrial genome, present as distinct subclades in the Spec and Elliot Uplands and Hinchinbrook Island. Interestingly, the same phenomenon may have occurred in a rainforest frog, *Litoria nannotis*, distributed across the same region that shows a similar pattern of cytonuclear discordance (Bell *et al*. 2012).

Why is introgression heterogeneous?

If this is the history explaining the pattern of genealogical discordance in this system, why is the pattern of discordance marker-specific? Heterogeneous introgression, particularly when the markers compared are cytoplasmic and nuclear, has several possible root causes: neutral and stochastic effects, selection or sex-biased processes. First, heterogeneous introgression can arise because of stochastic, neutral effects (Kuo & Avise 2005), like differences in drift among markers or the confounding effects of introgression and demography. Researchers have explained cytonuclear genealogical discordance by invoking the differential rates of drift and fixation in cytoplasmic and nuclear markers (Wilson & Bernatchez 1998). However, the balance between migration and drift is the same for both genomes because the higher drift in the mitochondrial genome is counteracted by a lower effective number of migrants (Wright 1951). Thus, drift is unlikely to explain this pattern. Yet, introgression does introduce stochasticity (Kuo & Avise 2005), and as the mitochondrial genome is just one marker, it might capture an extreme end of this variance. But, as our ABC-based demographic reconstructions suggest, such effects are still unlikely to lead to the sort of discordance seen here. Further, stochastic effects can be compounded by demographic events, such as changes in ranges or population growth (Ruedi *et al.* 1997; Currat *et al.* 2008). In particular, in the model of 'allele surfing', alleles from a resident population introgress readily into low-density populations at the edge of an expanding population (Currat *et al*. 2008). Loci that experience higher levels of drift—here, loci linked to the low-dispersing sex like the mitochondrial genome—are more likely to introgress quickly (Petit & Excoffier 2009; Melo-Ferreira *et al*. 2011). However, this work also predicts that even with low rates of admixture, complete replacement is expected at all loci, whether nuclear or mitochondrial (Currat *et al*. 2008). As such, we should see swamping of variation in the expanding population at both the nuclear and mitochondrial genomes. As we do not see this, we think allele surfing is unlikely to explain the patterns we see here.

Indeed, as shown by our demographic reconstruction, in order for introgression to commonly lead to strong genealogical discordance, gene flow levels must differ significantly between the two marker types—and selection and sex-biased processes are two factors that can lead to differential gene flow. Here, either a situation in which gene flow is elevated at the mitochondrial genome (model 5, Fig. 2B) or in which gene flow is depressed at the mitochondrial genome could explain the pattern we see (model 6; Fig. 2C). These two situations have the same geography, but the patterns of introgression differ and they reflect two very different biological realities. However, the genetic patterns, particularly when introgression is historical and introgressed alleles have been fixed, are hard to distinguish, and based on our demographic reconstructions, we cannot fully reject one or the other.

With respect to selection, rates of introgression and fixation are determined largely by the balance between migration and selection (Slatkin 1973). Selection can lead to loci having more limited or increased introgression compared with the background rate. Negative selection can take several forms—for example, the mitochondrial genome could be adapted to local bioclimatic niches and thus would introgress less readily (Ballard & Whitlock 2004; Cheviron & Brumfield 2009). This seems unlikely here, as the Southern and Central lineages do not occupy markedly different bioclimatic space (Fig. S6). Alternatively, cytonuclear incompatibilities (e.g. *Tigriopus californicus*; Ellison & Burton 2008) could limit introgression at the mitochondrial genome and a subset of the nuclear genome, while the rest of the nuclear genome would introgress freely (Takahata & Slatkin 1984). Cytonuclear incompatibilities can evolve owing to selection or owing to drift (Gavrilets 2003), and they are a plausible explanation for this discordance. In other studies, selective sweeps have lead to the cytoplasmic genome introgressing rapidly into other lineages (Gompert *et al.* 2008). However, the Southern mitochondrial lineage is deeply structured among southern isolates and, thus, does not have the typical signature of a selective sweep (i.e. low but rare diversity among haplotypes).

Finally, sex-specific processes could also explain these patterns. For example, if there is female philopatry, then the maternally inherited mitochondrial genome will experience less gene flow, which could lead to less introgression upon secondary contact (but see Petit & Excoffier 2009 for an alternative perspective). This is a common explanation for discrepancies between the mitochondrial and nuclear genomes (Berthier *et al.* 2006), and as we have evidence for male-biased dispersal in related species (Dubey & Shine 2010), this might explain our results as well. Further, patterns of

mating, whether owing to differences in population density (Hubbs 1955) or active mate choice, can lead to rapid introgression of the maternal mitochondrial genome across lineage boundaries (Chan & Levin 2005). We have no evidence either against or in support of this hypothesis, but it certainly could be a factor. In sum, it appears that the genealogical discordance we see in this system results from historical introgression, and it is quite possible this introgression acted in concert with either selection or sex-biased processes. Knowing how these selective or sex-biased processes are contributing and interacting to lead to this pattern cannot be determined from genetic data alone, but it is amenable to future study through field observation and experiments.

Significance

Examples of genealogical discordance are many, and cytonuclear discordance accounts for a significant number of these cases (Chan & Levin 2005; Currat *et al.* 2008). Stochastic effects, whether arising in the presence or absence of gene flow, could certainly explain many instances of discordance. However, many other cases, like ours, likely arise because of the unique biology of the mitochondria—i.e. selection limits introgression because of cytonuclear incompatibilities (Ellison & Burton 2008) or mate choice patterns promote introgression in sex-linked markers (Hubbs 1955; Chan & Levin 2005). Further, most examples of differential introgression across mitochondrial and nuclear genomes are cases where the introgressed marker has not yet reached fixation (Leaché 2009; Melo-Ferreira *et al.* 2011). In this respect, this system stands alongside a few other cases, most notably the arctic charr, North American chipmunks, polar bear and two species of temperate hares (Wilson & Bernatchez 1998; Reid *et al.* 2012; Hailer *et al.* 2012; Melo-Ferreira *et al.* 2012) in which distinct evolutionary units have seemingly become fixed for discordant mitochondrial DNA. Both this system and these other species share a history of small populations; the increased efficacy of drift in such populations might have quickened replacement of introgressing alleles. In other systems, where introgression often occurs between large populations that are expanding from glacial refugia (Hewitt 2011), insufficient time might have passed to allow introgressed alleles to reach fixation.

Finally, our and others' results suggest that introgression is likely manifest in the natural world. Because very little hybridization, or mating between different lineages, is necessary to spur introgression (Anderson 1949), the frequency of introgression might say very little about the frequency of hybridization. However, it does hint at interesting population histories of infrequent but dynamic changing hybridization, and it certainly suggests that introgression could be a pervasive and powerful force shaping the diversity of species and their populations (Mallet 2005). Whether introgression erodes genetic divergence as we see here or leads to adaptive change (Anderson 1949; Song *et al.* 2011), it is an undeniably important evolutionary process.

Acknowledgements

We gratefully acknowledge all those in the Moritz and Williams laboratories who collected samples for this work, in particular A. Moussalli, S. Williams and C. Hoskin. Assistance in the lab was provided by K. Whittaker and A. Xu and advice by members of the Moritz Lab, M. Hickerson, J. McEntee and M. Tonione. We thank A. Davis Rabosky, M. Hickerson, A. Leaché, J. McGuire, R. Pereira, M. Slatkin and two anonymous reviewers for thoughtful comments on earlier versions of this manuscript. Financial support was provided by the Museum of Vertebrate Zoology Koford Fund, a NSF Graduate Research Fellowship to SS, and NSF DEB0416250 to CM. The Texas Advanced Computing Center (TACC) at The University of Texas at Austin provided grid resources that contributed to the research results reported within this study.

References

Aljanabi S, Martinez, I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.

Anderson E (1949) Introgressive Hybridization. Wiley, New York.

Arevalo E, Davis S, and Sites J (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (*Phyrnosomatidae*) in Central Mexico. *Systematic Biology*, **43**, 387–418.

Ballard J, Whitlock M (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Beaumont M (2008) Joint determination of topology, divergence time and immigration in population tree. In: Simulation, Genetics and Human Prehistory (eds Matsumura S, Forster P), pp. 135–154. University of Cambridge Press, Cambridge, UK.

Beaumont M (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Bell R, Parra J, Tonione M *et al.* (2010) Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology*, **19**, 2531–2544.

Bell R, Mackenzie J, Hickerson M *et al.* (2012) Comparative multilocus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **279**, 991–999.

Berthier P, Excoffier L, Ruedi M (2006) Recurrent replacement of mtDNA and cryptic hybridization between two sibling species *Myotis myotis* and *Myotis blythii*. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **273**, 3101–3109.

Brandley M, Wang Y, Guo X *et al.* (2011) Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology*, **60**, 3–15.

Chan K, Levin S (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Cheviron Z, Brumfield R (2009) Migration-selection balance and local adaptation of mitochondrial haplotypes in Rufous-collared Sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*, **63**, 1593–1605.

Couper P, Keim L (1998) Two new species of *Saproscincus* (*Reptilia: Scincidae*) from Queensland. *Memoirs of the Queensland Museum*, **42**, 465–473.

Csilléry K, Francois O, Blum M (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution*, **3**, 475–479.

Currat M, Ruedi M, Petit R, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**, 1908–1920.

Degnan J, Rosenberg N (2009) Gene tree discordance, phylogenetic inference, and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**, 332–340.

Dolman G, Moritz C (2006) A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). *Evolution*, **60**, 573–582.

Dolman G, Phillips B (2004) Single copy nuclear DNA markers characterized for comparative phylogeography in Australian Wet Tropics rainforest skinks. *Molecular Ecology Notes*, **4**, 185–187.

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

Drummond A, Ashton B, Buxton S *et al.* (2010) Geneious v5.3. http://www.geneious.com.

Dubey S, Shine R (2010) Restricted dispersal and genetic diversity in populations of an endangered montane lizard (*Eulamprus leuraensis*, *Scincidae*). *Molecular Ecology*, **19**, 886–897.

Earl D (2011) Structure Harvester v0.6.6. http://taylor0.biology.ucla.edu/struct_harvest.

Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–97.

Edwards S (2009) Is a new and general theory of molecular systematics emerging? *Evolution* **63,** 1–19.

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

Ellison C, Burton R (2008) Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 15831–15836.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fagundes N, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17614–17619.

Gavrilets S (2003) Models of speciation: what have we learned in 40 years? *Evolution* **57**, 2197–2215.

Gelman A, Carlin J, Stern H, Rubin D (2004) Bayesian Data Analysis. Chapman & Hall, CRC, London, UK.

Gompert Z, Forister M, Fordyce J, Nice C (2008) Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, **17**, 5231–5244.

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

Hailer F, Kutschera V, Hallstrom B *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science*, **336**, 344–347.

Hamilton G, Stoneking M, Excoffier L (2005) Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7476–7480.

Hewitt G (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica*, **139**, 617–638.

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

Hubbs C (1955) Hybridization between fish species in nature. *Systematic Zoology*, **4**, 1–20.

Hudson D, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Huelsenbeck J, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

Jakobsson M, Rosenberg N (2007) CLUMPP: a cluster matching and permutation program for dealing with multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jennings W, Edwards S (2005) Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution*, **59**, 2033–2047.

Jesus F, Wilkins J, Wakeley J (2006) Expected coalescence times and segregating sites in a model of glacial cycles. *Genetics and Molecular Research*, **5**, 466–474.

Joly S, Bruneau A (2006) Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology*, **55**, 623–636.

Kingman J (1982) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Kuo C, Avise J (2005) Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. *Genetica*, **124**, 179–186.

Leaché A (2009) Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). *Systematic Biology*, **58**, 547–559.

Lenormand T, Roze D, Rousset F (2009) Stochasticity in evolution. *Trends in Ecology and Evolution*, **24**, 157–165.

Li J, Yeung C, Tsai P *et al.* (2010) Rejecting strictly allopatric speciation on a continental island: prolonged post-divergence gene flow between Taiwan (*Leucodioptron taewanus*, *Passeriformes*, *Timaliidae*) and Chinese (*L. canorum canorum*) hwameis. *Molecular Ecology*, **19**, 494–507.

Maddison W (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–526.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**, 229–237.

Marin J, Pillai N, Robert C, Rousseau J (2011) Relevant statistics for Bayesian model choice. Available on arXiv: 1110.4700v1.(In Review)

Maroja L, Andres J, Harrison R (2009) Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*, **63**, 2999–3015.

McGuire J, Linkem C, Koo M *et al.* (2007) Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of Crotaphytid lizards. *Evolution*, **61**, 2878–2897.

Melo-Ferreira J, Alves P, Rocha J, Ferrand N, Boursot P (2011) Interspecific X-chromosome and mitochondrial DNA introgression in the Iberian hare: selection or allele surfing. *Evolution*, **65**, 1956–1968.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves P, Farelo L, Alves P (2012) Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Systematic Biology*, **61**, 367–381.

Moritz C, Hoskin C, MacKenzie J *et al.* (2009) Identification and dynamics of a cryptic suture zone in a tropical rainforest. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **276**, 1235–1244.

Moussalli A, Moritz C, Williams S, Carnaval A (2009) Variable responses of skinks to a common history of rainforest fluctuation: concordance between phylogeography and palaeo-distribution models. *Molecular Ecology*, **18**, 483–499.

Near T, Bossu C, Bradburd G *et al.* (2011) Phylogeny and temporal diversification of darters (*Percidae: Etheostomatinae*). *Systematic Biology*, **60**, 565–595.

Nylander J (2004) Mrmodeltest v2. program distributed by the author. *Evolutionary Biology Center Uppsala University*, **2**, 1–2.

Petit R, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **145**, 945–959.

Reid N, Demboski JR, Sullivan J (2012) Phylogeny estimation of the radiation of western American chipmunk (*Tamias*) in the face of introgression using reproductive protein genes. *Systematic Biology*, **61**, 44–62.

Robert C, Corneut J, Marin J, Pillai N (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 15112–15117.

Rosenberg N (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Resources*, **4**, 137–138.

Rosenblum E, Hoekstra H, Nachman M (2004) Adaptive reptile color variation and the evolution of the *MC1R* gene. *Evolution*, **58**, 1794–1808.

Ruedi M, Smith M, Patton J (1997) Phylogenetic evidence of mitochondrial DNA introgression among pocket gophers in New Mexico, family: *Geomyidae*. *Molecular Ecology*, **6**, 453–462.

Saint K, Austin C, Donnellan S, Hutchinson M (1998) C-mos, a nuclear marker useful for squamate phylogenetic analysis. *Molecular Phylogenetics and Evolution*, **10**, 259–263.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics*, **75**, 733–756.

Slatkin M, Pollack J (2008) Subdivision in an ancestral species creates asymmetry in gene trees. *Molecular Biology and Evolution*, **25**, 2241–2246.

Song Y, Endepois S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Current Biology*, **21**, 1296–1301.

Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

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

Swofford D (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Takahata N, Slatkin M (1984) Mitochondrial gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 1764–1767.

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

Thornton K, Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics*, **172**, 1607–1619.

VanDerWal J, Shoo L, Williams S (2009) New approaches to understanding late Quaternary climate fluctuations and refugial dynamics in Australian tropical rain forests. *Journal of Biogeography*, **36**, 291–301.

Wakeley J (2008) Coalescent Theory: An Introduction. Roberts and Company, Greenwood Village, Colorado.

Williams S, Pearson R, Walsh J (1996) Distributions and biodiversity of the terrestrial vertebrates of Australia's Wet Tropics: a review of current knowledge. *Pacific Conservation Biology*, **2**, 327–362.

Wilson C, Bernatchez L (1998) The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology*, **7**, 127–132.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Yeung C, Tsai P, Chesser RT *et al.* (2011) Testing founder effect speciation: Divergence population genetics of the spoonbills *Platalea regia* and *Pl. minor* (Threskiornithidae, Aves). *Molecular Biology Evolution*, **28**, 473–482.

S.S. is a graduate student at VC Berkeley, this work is part of her thesis investigating the causes and unsequenses of introgression in an Australian suture zone. C.M. is an evolutionary biologist interested in explaining patterns of biodiversity through both space and time.

Data accessibility

Data are available at the following locations:

1 DNA sequences are available on GenBank, accessions JX313797–JX315247.

2 Information on PCR conditions for loci and sample locations are uploaded as online supporting information.

3 Final alignments and scripts used for simulations are available on DRYAD, entry doi:10.5061/dryad.dn0m4.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Gene trees for eight nuclear genes for *Saproscincus basiliscus* and *S. lewisi* based on individual haplotypes, as inferred by maximum likelihood in RAxML. Colour scheme follows that used in Fig. 1.

Fig. S2 Expected distribution of mito-nuclear divergence ratios for all of the modelled scenarios across the complete parameter space, (A) before fitting and (B) after fitting. The mito-nuclear divergence ratio found in this study is outlined in darker grey.

Fig. S3 Prior (shown by dotted line) and posterior (shown by bold black line) probability distributions for parameters of the most likely inferred model (model 5), in which there is pulsed introgression with more mitochondrial than nuclear gene flow.

Fig. S4 Results from 100 pseudo-observed data sets, showing the frequency of mis-classification among the models simulated for the approximate Bayesian computation (ABC) analysis.

Fig. S5 Posterior predictive results for all six models across all seven summary statistics. Dashed black lines reflect true value of summary statistic for the empirical data. Some graphs cropped for ease of visibility.

Fig. S6 PCA of climatic variables (Bioclim) grouped by mitochondrial lineage; PCA performed using prcomp in R. Colour scheme follows that used in Fig. 1.

Table S1 Data on sampled individuals.

Table S2 Loci used in this study, including their associated information.

Table S3 Prior distributions for parameters used in simulating data sets for the approximate Bayesian computation (ABC) analysis.

Table S4 Type I and Type II errors for model mis-classification based on pseudo-observed data set analysis.

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