

1-1-2015

## Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: Tribolonotus)

Eric N. Rittmeyer  
*Louisiana State University*

Christopher C. Austin  
*Louisiana State University*

Follow this and additional works at: [https://repository.lsu.edu/biosci\\_pubs](https://repository.lsu.edu/biosci_pubs)

---

### Recommended Citation

Rittmeyer, E., & Austin, C. (2015). Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: Tribolonotus). *Molecular Ecology*, 24 (2), 466-483.  
<https://doi.org/10.1111/mec.13030>

This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact [ir@lsu.edu](mailto:ir@lsu.edu).

# Combined next-generation sequencing and morphology reveal fine-scale speciation in Crocodile Skinks (Squamata: Scincidae: *Tribolonotus*)

ERIC N. RITTMAYER<sup>1</sup> and CHRISTOPHER C. AUSTIN

Department of Biological Sciences, Museum of Natural Science, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA

## Abstract

Next-generation sequencing has vast potential to revolutionize the fields of phylogenetics and population genetics through its ability to collect genomic scale data sets of thousands of orthologous loci. Despite this potential, other types of data (e.g. morphology, ecology) remain important, particularly for studies endeavouring to delimit species. Here, we integrate next-generation sequencing with morphology to examine divergence between populations of *Tribolonotus pseudoponceleti* on the islands of Buka and Bougainville in the Solomon Archipelago. We used the Ion Torrent PGM to collect over 648 Mbp of sequence data for 12 samples, representing 1526 loci recovered from all samples, and 3342 were recovered from at least six samples. Genetic structure analyses strongly support the distinctiveness of these two populations, and Bayes factor delimitations decisively select speciation between Buka and Bougainville. Principal components and discriminant function analyses reveal concordant morphological divergence. Finally, demographic analyses via diffusion approximation and approximate Bayesian computation prefer a complex model of mid-Pleistocene divergence with migration, and a later decrease or cessation of migration and population size shift, suggesting a scenario in which migration was enabled by Pleistocene merging of these two islands, and limited when isolated by higher sea levels. Further analysis of four Sanger sequenced loci in IMA2 had limited power to distinguish among models including and excluding migration, but resulted in similar population size and divergence time estimates, although with much broader confidence intervals. This study represents a framework for how next-generation sequencing and morphological data can be combined and leveraged towards validating putative species and testing demographic scenarios for speciation.

**Keywords:** approximate Bayesian computation, genomics, Solomon Archipelago, species delimitation, species validation

Received 23 April 2013; revision received 17 November 2014; accepted 28 November 2014

## Introduction

With the rapidly expanding availability of next-generation sequencing (NGS) technologies capable of collecting genomic data sets of a scale previously restricted to

model systems, and the increasing availability of computation tools capable of analysing the resultant data sets, these vastly superior new technologies have enormous potential to revolutionize the studies of phylogenetics and population genetics through enabling researchers to obtain far more independently evolving loci for a much lower cost per base pair of sequence data (Metzker 2010; McCormack *et al.* 2012, 2013; Wagner *et al.* 2013). These massively multilocus data sets can be leveraged to elucidate evolutionary patterns

Correspondence: Eric N. Rittmeyer, Fax: +61 2 6125 5573;

E-mail: eric.rittmeyer@anu.edu.au

<sup>1</sup>Current address: Research School of Biology, The Australian National University, Canberra, ACT 2601, Australia

and processes with unprecedented power and precision and thus to transform these fields. Several studies have shown that many evolutionarily important parameters, such as population size, divergence times and migration rates, require data sets with large numbers of independent loci, of a scale previously restricted to model systems (Carling & Brumfield 2007). The increasing application and availability of NGS technologies is already enabling researchers to disentangle the evolutionary histories of nonmodel systems with increasing accuracy and robustness, bettering our understanding of the processes responsible for the generation of biodiversity.

Despite the ability of NGS to generate large genome-wide DNA sequence data sets, other types of data, such as morphological, ecological or acoustic, remain critically important, particularly in studies endeavouring to clarify systematics and species limits in complexes of closely related species (O'Meara 2010; Yang & Rannala 2010; Rittmeyer & Austin 2012). Although species are a fundamental unit in biology, vital to a wide variety of disciplines and conservation efforts (Bickford *et al.* 2007), much contention remains over species concepts and how to identify and delimit species (Coyne & Orr 2004; De Queiroz 2005, 2007). This controversy is likely due largely to the complex and gradual nature of speciation and to differences in the processes driving speciation among systems (Coyne & Orr 2004). Dependent on the specific evolutionary forces acting on a system, morphological data can reveal extremely recent divergences that may be difficult to distinguish with molecular data (Lance *et al.* 2008; Rheindt *et al.* 2011; McCormack *et al.* 2012). Alternatively, molecular analyses frequently reveal previously overlooked lineages that are deeply divergent yet morphologically similar (Bickford *et al.* 2007; Arbogast & Kenagy 2008), although subsequent detailed morphological analyses often reveal subtle characters that can distinguish among these lineages identified using molecular data (Burbrink 2001; Pyron & Burbrink 2009). Identifying these characters facilitates field studies and conservation efforts through enabling researchers to identify museum vouchers and specimens in the field without necessitating laboratory analysis, thus enhancing the digitization and dissemination of biodiversity records to scientists and the general public.

Due to variation in the process of speciation, arguably the most generally applicable species concept is the general lineage species concept (but see Baum 2009; Hausdorf 2011; Naomi 2011), which defines species as independently evolving lineages (Baum & Shaw 1995; De Queiroz 2005, 2007; Baum 2009). The general lineage species concept argues that rather than defining species, other species concepts describe characteristics of species that evolve as these lineages diverge. In some situations

involving recently divergent lineages, multiple sources of data may be critical to the accurate delimitation of species, whereas in other, more diverged systems, fewer types of data may be necessary. Regardless of the level of divergence, through combining genomic data collected via NGS with other data sets, the species limits, as well as the evolutionary patterns and processes responsible for their generation, can be elucidated more robustly than either data set alone would accomplish. Here, we combine morphological data with a genomic scale data set collected via NGS to examine the divergence between Crocodile Skinks (Squamata: Scincidae: *Tribolonotus*) on the islands of Buka and Bougainville.

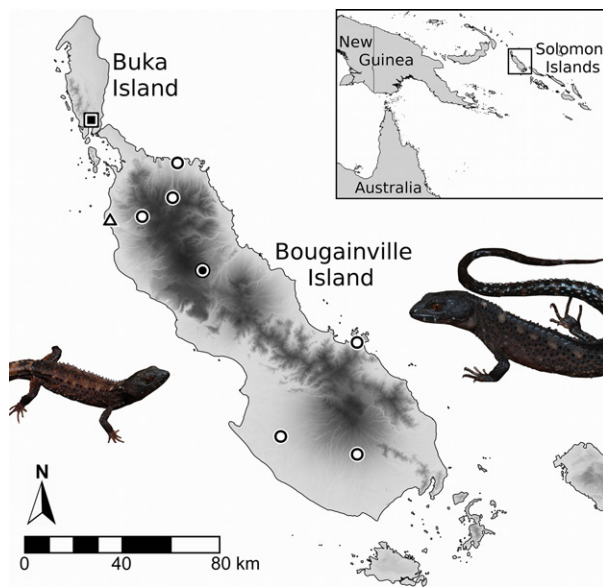
The Crocodile Skinks of the genus *Tribolonotus* are a group of eight species distributed throughout northern New Guinea and the northern Melanesian islands of the Admiralty, Bismarck and Solomon Archipelagos (Cogger 1972; McCoy 2006). The genus is united by the presence of two peculiar synapomorphies: abdominal glands, and palmar and plantar pores (Zweifel 1966; Greer & Parker 1968; McCoy 2006). *Tribolonotus* is also unusual among scincid lizards in having strongly keeled or spinose scales and, in at least two species (*T. gracilis*, *T. ponceleti*), the ability to vocalize (Hartdegen *et al.* 2001; McCoy 2006), an ability known from only one other scincid genus (*Nannoscincus*, Bauer *et al.* 2004). In a recent phylogenetic study of the genus, Austin *et al.* (2010) found that *T. pseudoponceleti* from the islands of Buka and Bougainville in the northwestern Solomon Archipelago are reciprocally monophyletic and deeply divergent (4.3% divergent at the mitochondrial cytochrome B and NADH dehydrogenase subunit 2 loci). This divergence is surprising given the geographic proximity and geologic history of the islands: Buka and Bougainville are currently separated by the Buka Passage, a narrow channel only approximately 300 m in width. Further, the islands were historically connected multiple times in periods of lower sea levels during Pleistocene glaciations, including as recently as the last glacial maximum less than 20 000 years ago, forming a larger Greater Bougainville Island, along with the islands of Choiseul and Isabel further to the southeast in the Solomon Archipelago (Chappel & Shackleton 1986; Mayr & Diamond 2001). While Austin *et al.* provided the first molecular evidence of a divergence between the populations of *T. pseudoponceleti* from Bougainville and Buka islands, this study focused on the deeper level phylogeny and evolution of the genus *Tribolonotus*, and did not examine this divergence in any detail, test for any morphological differences between these populations or test whether this divergence may represent species-level differentiation. Here, we combine morphological data with genomic data collected via next-generation sequencing (NGS) to examine

this divergence in detail and (i) test whether the Buka and Bougainville populations represent distinct species, (ii) examine the demographic history and potential drivers of divergence and (iii) test for morphological divergence and diagnosability of the two populations. We also reanalyze the Sanger data of Austin *et al.* (2010) to compare demographic parameter estimates obtained using a small number of long loci to those obtained from the genomic data. While *T. pseudoponceleti* is also known from the northwestern tip of Choiseul Island, Solomon Archipelago, no genetic samples are available from this island; thus, we here focus exclusively on examining the divergence between the populations on the islands of Buka and Bougainville.

## Methods

### Morphological data and analyses

To test for morphological divergence between the Buka and Bougainville populations of *Tribolonotus pseudoponceleti*, we examined a total of 39 specimens from Buka and 115 specimens from 11 populations spanning Bougainville (1–42 specimens per population, mean 10.5), including the holotype from Kunua, Bougainville (Fig. 1, Table S1, Supporting information). For each



**Fig. 1** Sampling localities of *Tribolonotus pseudoponceleti* specimens included in this study. Filled symbols represent specimens included in both molecular and morphological analyses; open symbols represent specimens included only in morphological analyses. Squares represent specimens from Buka, circles represent specimens from Bougainville, and the type locality of *T. pseudoponceleti* (Kunua, Bougainville) is shown as a triangle. Inset images show representatives of the Buka (left) and Bougainville populations (right).

specimen, eight mensural and eight meristic characters were scored. Mensural characters included first, third and fifth finger lengths, first, fourth and fifth toe lengths, forelimb length and hindlimb length. Meristic characters included first, third and fifth finger subdigital lamellae, first, fourth and fifth toe subdigital lamellae, number of enlarged paravertebral scale rows and number of ventral scale rows.

Principal component analyses (PCAs) were implemented on the mensural and meristic data sets in R ver. 2.15.1 to test for morphological divergences between Buka and Bougainville populations. Prior to PCAs, all mensural characters were scaled by snout-vent length (SVL) to correct for differences in body size among individuals. To test for morphological diagnosability, discriminant function analyses (DFAs) were also implemented on the mensural and meristic data sets in R using the MASS package (Venables & Ripley 2002). As in PCAs, all mensural characters were scaled by SVL prior to DFAs.

### Next-generation sequencing

One of the biggest challenges with the application of NGS technology to population genetics studies is the need to reduce the genome such that numerous orthologous loci can be sequenced for many individuals (Hird *et al.* 2011; McCormack *et al.* 2012). This can be particularly challenging in organisms with particularly large genomes, such as salamanders (McCormack *et al.* 2013; Gregory 2014); while the genome size of *Tribolonotus* is unknown, a related species, *Tiliqua scincoides*, has a genome of 1.780 Gb (De Smet 1981; MacCulloch *et al.* 1996; Gregory 2014), and other scincid lizards have genomes ranging from 1.027 Gb in *Chalcides mionecton* (De Smet 1981; Gregory 2014) to 3.130 Gb in *Chalcides ocellatus* (Capriglione *et al.* 1987; Gregory 2014). To accomplish genome reduction, we applied a double-digest approach, similar to amplified fragment length polymorphism (AFLP) analysis (Vos *et al.* 1995; McCormack *et al.* 2012). Whole genomic DNA was extracted from liver taken from 12 individuals, including five individuals from Bougainville, five individuals from Buka (Fig. 1, Table S1, Supporting information) and two individuals of the sister taxon, *T. ponceleti*, using a Qiagen DNeasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, USA) as per manufacturer's instructions. Library preparation protocols largely followed McCormack *et al.* (2012), but several modifications were made. First, digestion and ligation steps were separated: approximately 250 ng of DNA extracts were completely digested with EcoRI and MseI restriction enzymes (New England Biolabs, Ipswich, MA, USA) with 5 units of EcoRI and 1 unit of MseI in 1X T4 DNA Ligase

buffer (New England Biolabs), 0.061 mg/mL BSA and 0.056 M NaCl by incubation for 4 h at 37 °C, followed by 20 min incubation at 65 °C to denature the restriction enzymes. AFLP adapters were then ligated to the DNA digests by adding 0.091  $\mu$ M of each adapter (Table S2, Supporting information) and 20 units of T4 DNA ligase (New England Biolabs) and incubating at 16 °C for 4 h. Second, both amplification steps were conducted using Platinum PCR SuperMix High Fidelity (Invitrogen, Carlsbad, CA, USA) as per manufacturer's instructions using the MseI-PreAmp primer and a biotinylated EcoRI-PreAmp primer (Table S2, Supporting information). Third, due to the shorter read lengths of the Ion Torrent PGM, amplicons in the 100–160 bp range were excised from the gel and extracted using a Qiagen Qiaquick gel extraction kit (Qiagen, Inc., Valencia, CA, USA) as per manufacturer's instructions, and gel extracts were purified to isolate biotinylated fragments (i.e. those with EcoRI cut sites) using Dynabeads MyOne Streptavidin C1 (Invitrogen, Carlsbad, CA, USA) as per manufacturer's instructions. Libraries for each sample were then purified using Agencourt AMPure XP beads (Agencourt Bioscience, Beverly, MA, USA) as per manufacturer's instructions, quantified at two or more dilution points per sample, and six samples were pooled into each of two equimolar libraries. Emulsion PCR was then applied on each library to amplify library DNA onto Ion Particle Spheres using an Ion OneTouch Template Kit as per manufacturer's instructions, enriched and sequenced using 316 chips on an Ion Torrent PGM. All primer and adapter sequences are provided in the Appendix (Table S1, Supporting information).

### *Locus and allele calling*

Sequence reads from each 316 chip were initially processed using the Ribosomal Database Project Pyrosequencing Initial Process (Cole *et al.* 2009) to sort reads by barcode and filter out reads with a length below 50 bp or with a minimum quality score of below Q20. The PRGmatic v1.6 pipeline (Hird *et al.* 2011) was then used to identify orthologous DNA sequences (hereafter referred to as 'loci') and call alleles. A minimum threshold of 10 $\times$  coverage was applied to call high confidence alleles, and the PRGmatic default setting of 90% identity was used to call loci. Minimum coverage for calling consensus sequences in an individual was set to 10 $\times$ , and the minimum coverage for calling a single nucleotide polymorphism (SNP) in an individual was 5 $\times$ . Muscle ver. 3.8.31 (Edgar 2004) was then used to align the identified loci. As the threshold of 90% identity for calling loci may result in paralogous loci being combined as a single locus, we examined all loci for

multiple (>2 bp) SNP calls at a single site for an individual. Any loci where at least one individual had >2 bp (including gaps) at a single position in either >5% of the total reads for that locus or in >3 reads total were discarded as potentially paralogous.

### *Genetic structure analyses*

We examined the structure of the populations using genetic clustering algorithms implemented in Structure (Pritchard *et al.* 2000; Falush *et al.* 2003) and Structurama (Huelsenbeck & Andolfatto 2007; Huelsenbeck *et al.* 2011). While both programs apply the same algorithm to attempt to maximize Hardy–Weinberg equilibrium and minimize linkage disequilibrium by clustering the samples into populations, Structurama has the added benefit of implementing a Dirichlet process prior to estimate the number of clusters and the assignments of individuals to clusters simultaneously (Huelsenbeck & Andolfatto 2007). However, both Structure and Structurama assume that loci are unlinked allelic markers and thus require the coding of each locus as alleles or the calling of a single SNP from each locus obtained from the NGS runs. We prefer the method of coding the alleles at each locus as integers because it results in less loss of data than calling a single SNP from each locus, although both approaches result in the loss of genetic distance information among alleles. We used a custom python script, seq2struct, to collapse each NGS locus to alleles coded as integers and to prepare the input files for both Structure and Structurama. This and all subsequent custom scripts are available via the Dryad repository, doi: 10.5061/dryad.87550. Three sets of analyses were run, all of which included all 12 individuals, but varied in their completeness and the number of loci: one including only those loci sequenced in all 12 individuals (i.e. no missing data), one including all loci sequenced in at least nine of the 12 individuals (i.e. including all 12 individuals and all loci included in the previous data set, but allowing for some missing data to include loci sequenced in nine, 10 or 11 individuals) and a third including all loci sequenced in at least six of the 12 individuals (i.e. including all 12 individuals and all loci included in the previous two data sets, but allowing for missing data to include loci sequenced in six, seven or eight individuals).

Structure analyses were implemented in Structure ver. 2.3.4 and consisted of 20 replicates at each value for  $K$  (the number of clusters) from one to ten. Each run consisted of a burnin of  $1 \times 10^5$  iterations, followed by a sampling period of  $1 \times 10^5$  iterations, sampling every 100 iterations, with correlated allele frequencies and all other parameters set to default values. To estimate the optimal number of clusters, the  $\Delta K$  statistic (Evanno

*et al.* 2005) was calculated using Structure Harvester (Earl & VonHoldt 2012), and the estimated  $\ln$  probability of the data,  $\ln \Pr(D|K)$ , was examined. Several additional, longer runs were also conducted with a burnin of  $5 \times 10^5$  iterations and a sampling period of  $1 \times 10^6$  iterations at a variety of values of  $K$ . These longer runs resulted in similar values of likelihood,  $\ln \Pr(D|K)$ , and cluster assignments as the shorter runs, thus the full 20 replicates for each value of  $K$  were only conducted for the shorter runs.

To further examine the population structure within *T. pseudoponceloti*, we used Structurama to implement the Dirichlet process prior in estimating the number of clusters and cluster membership. A variety of prior distributions on the  $\alpha$  parameter of the Dirichlet process prior, which controls the probability of two samples being assigned to the same cluster and thus impacts the number of clusters inferred, were tested, including an exponential distribution with a fixed prior mean number of populations, with the mean number of populations set to one through five and three gamma distributions:  $G\sim(1.0,1.0)$ ,  $G\sim(2.5,0.5)$  and  $G\sim(0.5,2.5)$ , selected based on a series of preliminary analyses and to provide a broad range of prior distributions on the number of populations. For all eight prior distributions on the  $\alpha$  parameter, analyses were conducted for each of three data sets, as in the Structure analyses: loci sequenced from all twelve samples, loci sequenced in at least nine samples and loci sequenced in at least six samples. Structurama analyses were run for  $1.5 \times 10^6$  iterations, sampling every 100 iterations, with the first 5000 samples discarded as burnin. All other priors were left at default values.

#### *Species delimitation using Bayes factors*

To test between the two competing species delimitation hypotheses (i.e. two species: *T. ponceloti*, Buka + Bougainville vs. three species: *ponceloti*, Buka, Bougainville) in a species coalescent framework, we used Bayes factors to compare the species trees estimated under each of these models in SNAPP (Bryant *et al.* 2012; Leaché *et al.* 2014). SNAPP requires biallelic SNP data and assumes that all SNPs are unlinked (Bryant *et al.* 2012); thus, we used a custom python script, biSNPcaller to select the first biallelic SNP from each locus. We tested the impact of a variety of prior distributions on each parameter, including Gamma (2, 2), Gamma (2, 10) and Gamma (2, 100) distributions on the mutation rate parameters, an improper infinite uniform (0,  $\infty$ ), and a uniform (0, 1e5) distribution on the Yule speciation rate, and a Gamma (2, 100), Gamma (2, 1000), Gamma (2, 10 000) on theta in a series of preliminary runs, selected based on previous studies (Leaché *et al.* 2014;

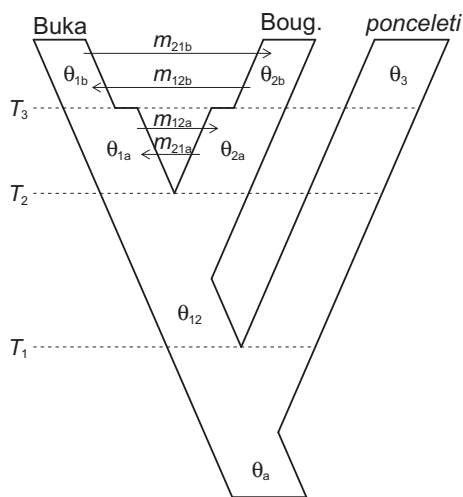
Rheindt *et al.* 2014) and to provide a range of possible prior distributions. Prior distributions on the mutation rate and Yule speciation rate parameters had no impact on the resultant posterior probability, likelihood, or parameter estimates (results not shown), so default settings we retained for each of these priors. Selection of prior distribution on theta, however, substantially impacted the results: smaller prior means resulted in smaller estimated population sizes and shallower divergences. Therefore, full analyses were run under each of the three aforementioned priors on theta. Similar to the Structurama analyses, SNAPP analyses were conducted on three data sets: one including only those loci sequenced in all 12 samples, one including all loci sequenced in at least nine of the 12 samples and a third including all loci sequenced in at least six of the 12 samples. Marginal likelihoods were estimated via path sampling (Lartillot & Philippe 2006) in BEAST ver. 2.1.0 (Bouckaert *et al.* 2013) with the SNAPP ver. 1.1.4 (Bryant *et al.* 2012) and BEASTii ver. 1.1.0 add-ons, which implements the proportionality constant correction described by Leaché *et al.* (2014) to make marginal likelihoods estimated under different species delimitation models comparable. Each path sampling analysis included 48 steps, each consisting of  $1 \times 10^5$  generations, the first 10% of which were discarded as burnin, which was sufficient to obtain suitable effective sample sizes (>200). Bayes factors were then calculated as  $2 \times \ln BF$ , where BF is the ratio of the marginal likelihoods for the two competing models (i.e. twice the difference in the log marginal likelihood), and evaluated following the framework of Kass & Raftery (1995). We subsequently estimated the species trees and relevant evolutionary parameters (effective population size, divergence times) in SNAPP (i.e. without path sampling) for each of the data sets and priors under the preferred species delimitation model. Analyses were run for  $2 \times 10^6$  iterations, sampling every 1000 iterations, the first 10% of which were discarded as burnin.

#### *Demographic model selection using genomic data*

The fitting of demographic models, such as isolation with migration (IM) class models, represents a powerful method for comparing among evolutionary hypotheses and thus for testing for the importance of various evolutionary parameters, such as migration or population size changes (Hey & Nielsen 2007; Carstens *et al.* 2009; Hey 2010; Sousa & Hey 2013). The size of data sets obtained via NGS makes full Bayesian implementation of demographic models computationally unfeasible, and the short read lengths obtained by some methods preclude gene genealogy-based analyses. However, the large number of loci obtained also enable comparisons

among more complex, and potentially more realistic, models, using alternative methods that are more computationally efficient, such as diffusion approximation of allele frequency spectra (Gutenkunst *et al.* 2009) or approximate Bayesian computation (Beaumont *et al.* 2002; Csilléry *et al.* 2010).

During Pleistocene glacial cycles, the islands of Buka and Bougainville repeatedly enlarged and merged during glaciations and shrunk and separated during interglacials (Chappell & Shackleton 1986; Shackleton 1987; Chappell *et al.* 1996). As the mitochondrial divergence suggests a more ancient divergence between the Buka and Bougainville populations than the last glacial maximum (LGM), we hypothesize that these populations may have diverged during an early interglacial, with some migration potentially occurring between the populations during later glacial cycles. Due to the more recent decrease in island area and isolation of the islands since the LGM, we further hypothesize that these populations may have experienced a shift in population size and migration rate later in their evolution. We test among nine models of the demographic history for these populations, varying from the simplest model of no divergence between Buka and Bougainville, to the most complex model, including a shift in the population sizes of, and migration rates between, Buka and Bougainville subsequent to their divergence (Fig. 2, Fig. S1, Supporting information). To compare among these models, we use two approaches: diffusion approximation of the allele frequency spectrum (AFS) and approximate Bayesian computation (ABC).



**Fig. 2** Schematic of the most complex demographic model (model 9) assessed in  $\partial a \partial i$  and ABC analyses. Note that population sizes following the shift at time  $T_3$  are not constrained to be smaller than population sizes prior to the shift; similarly, migration rates were allowed to increase or decrease at the shift. For details of additional demographic models tested, see Fig. S1 (Supporting information).

Diffusion approximation of the AFS was implemented in  $\partial a \partial i$  v. 1.6.3 (Gutenkunst *et al.* 2009).  $\partial a \partial i$  requires biallelic SNP data, thus, as for the SNAPP analyses, we used a custom python script, biSNPcaller, to identify the first biallelic SNP in each locus. Five search replicates were conducted for each model and compared using the Akaike information criterion (AIC). We then estimated confidence intervals from 100 nonparametric bootstrap replicates of the best fit model. All parameters are, by definition, positive; therefore, we used log-transformed parameters to estimate confidence intervals (i.e.  $e^{(\ln(\theta^*) \pm 1.96 \ln(\sigma))}$ , where  $\theta^*$  is the maximum-likelihood estimate, and  $\sigma$  is the standard deviation of the bootstrap replicates), following Gutenkunst *et al.* (2009).

ABC analyses consist of simulating a number of data sets under the models of interest and calculating a suite of summary statistics for each simulated data set (Beaumont *et al.* 2002; Csilléry *et al.* 2010). A set of simulated data sets with summary statistics most similar to the observed summary statistics are then accepted as an approximation of the Bayesian posterior distribution and can be used to assess model support and parameter values. To test among the nine demographic models using ABC, we used a custom python script to simulate two million data sets for each model (18 million total) and calculate a suite of summary statistics for each simulated data set using msABC (Pavlidis *et al.* 2010). The number of loci and the lengths of each locus in the simulated data sets were equivalent to the number and lengths of loci sequenced for all 12 individuals in the empirical data set, with ten alleles simulated for each of the Buka and Bougainville populations (i.e. five diploid individuals each) and four alleles for the sister taxon, *T. ponceleti* (i.e. two diploid individuals). Model parameters for the simulations were sampled from uniform prior distributions, with limits set based on the results of  $\partial a \partial i$  and SNAPP analyses, as well as through a series of preliminary simulations examining the similarity between the resultant simulated summary statistics vector and the observed summary statistics (Fig. S2, Supporting information). The upper limit of the divergence time parameters was set to twice the upper limit of the 95% confidence interval on divergence time from  $\partial a \partial i$  ( $1.21076 N_0$  generations). Using this upper limit, the divergence time of the Buka and Bougainville populations,  $t_1$ , was sampled from a uniform prior distribution of  $0-2.4215 N_0$  generations; the divergence time of *T. ponceleti* from *T. pseudoponceleti*,  $t_2$ , was sampled from a uniform prior distribution of  $t_1$  to  $2.4215 N_0$  generations, thus constraining the Buka and Bougainville populations to sister taxa (i.e. with a more recent divergence than *T. ponceleti* from *T. pseudoponceleti*). The time of the shift in population size and migration rate

within the Buka and Bougainville populations was sampled from a uniform prior distribution of 0 to  $t_1$ , thus constraining this to occur more recently than the divergence between these populations. The population size parameters,  $\theta$ , for each extant and ancestral population were sampled independently from a uniform prior distribution, with a lower limit of 0, and an upper limit equal to the maximum upper limit of the 95% highest posterior density (HPD) interval on theta from the SNAPP analyses (0.012013 in units of  $4N_0\mu$ ). Due to the breadth of the confidence intervals on the migration rate parameters in  $\hat{\alpha}\hat{\delta}i$ , we set the upper limit on these parameters to four times the largest maximum-likelihood estimate (7.8869 in units of  $4N_0m$ ), as this resulted in a more reasonable prior distribution that still included the entirety of the confidence intervals for three of the four migration rate parameters, and much of the confidence interval for the fourth parameter. Migration parameters were sampled from uniform prior distributions of 0–31.5478, in units of  $4N_0m$ , where  $N_0$  is the ancestral population size, and  $m$  is the proportion of the population composed of new migrants each generation. A suite of 50 summary statistics were calculated for the empirical and each simulated data set using msABC including the mean and variance across all loci for the following: number of segregating sites,  $\pi$ , and  $\theta_w$  for the complete data set and for each population, total  $F_{st}$  and, pairwise across all three pairs of populations (Buka + Bougainville, Buka + *T. ponceleti*, Bougainville + *T. ponceleti*), shared alleles, private alleles, fixed differences and  $F_{st}$ . Multinomial logistic regression was implemented using the abc package (Csilléry *et al.* 2012) in R to perform the model selection step with a tolerance of  $5.55 \times 10^{-5}$  to retain 1,000 simulated data sets. We then increased the number of simulations for the best fit model to  $5 \times 10^6$  total and applied local linear regression using the abc package in R to estimate parameters using a tolerance of  $2 \times 10^{-4}$  to retain 1000 simulated data sets. For each parameter, the geometric mean was estimated in R using the psych package (Revelle 2012), and the 95% HPD intervals were calculated in R using the coda package (Plummer *et al.* 2005).

The accuracy of ABC analyses is highly dependent on the correlation of the summary statistics used with the parameters of interest and their utility in discerning among the models being tested, as well as on the selection of appropriate prior distributions that result in simulated data sets with comparable summary statistics to the empirical values (Beaumont *et al.* 2002; Csilléry *et al.* 2010). To assess the appropriateness of the selected prior distributions in generating similar summary statistics, we applied PCA in R and plotted the first two principal components to visualize whether the empirical values fell within the cloud of values

from the simulated data sets. To validate the accuracy of ABC model selection and parameter estimation, 100 simulated data sets were selected at random for each of the nine models and used as pseudo-observed data for model selection and parameter estimation rejection steps.

No useful fossils are available to calibrate the demographic parameter estimates; therefore, we converted the raw parameters estimated from both the  $\hat{\alpha}\hat{\delta}i$  and ABC analyses to more biologically meaningful units using a relative substitution rate calculated from available estimates of mitochondrial substitution rates. Rates of mitochondrial substitution in related squamate reptiles range from approximately 0.01 to 0.02 substitutions/site/million years (Brown & Pestano 1998; Malhotra & Thorpe 2000; Carranza & Arnold 2003; Poulakakis *et al.* 2005; Jackson & Austin 2010, 2012); therefore, we use an estimated rate of 0.015 substitutions/site/million years to calculate a relative substitution rate for the NGS data. Nucleotide diversity,  $\pi$ , was calculated for each NGS locus and for the mitochondrial sequence data. We averaged  $\pi$  across all NGS loci and used the NGS  $\pi$ /mtDNA  $\pi$  ratio to scale the estimated mitochondrial substitution rate and calculate a relative substitution rate for the NGS loci, which was then used to calibrate the demographic parameters.

#### *Demographic modelling using Sanger data*

To compare the demographic parameter estimates to estimates based on a small number of longer loci collected via Sanger sequencing, we obtained two mitochondrial and three nuclear gene regions (i.e. four independent loci) previously sequenced for each of the 12 samples from GenBank (Austin *et al.* 2010; Table S3, Supporting information). Haplotypes at the nuclear loci were resolved using Phase ver. 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003). We then fit the standard isolation with migration models to this Sanger data set in IMA2 (Hey 2010), using the estimated mitochondrial substitution rate of 0.015 substitutions/site/million years. Due to the limitations of the IMA2 program, and the small size of the data set, the more complex models involving shifts in migration rate and population size subsequent to the divergence of the Buka and Bougainville could not be tested. Therefore, we focused only on the standard isolation with migration model and the related submodels involving limiting migration between these populations. After a series of shorter preliminary runs to assess the optimal prior distributions and heating strategy, full analyses were run using 10 chains with a geometric heating model (heating terms  $h_1 = 0.9$ ,  $h_2 = 0.85$ ) for 10 million iterations, sampled every 100 iterations, following a burnin period of 1 million



iterations. The resultant 100 000 posterior samples were then used in the 'load genealogies mode' to calculate the maximized posterior density function ( $-\log(P)$ ) for the full isolation with migration model and four sub-models: symmetrical migration, unidirectional migration (Buka to Bougainville and Bougainville to Buka) and no migration (i.e. pure isolation). The AIC and related information theoretic statistics ( $\Delta AIC$ , Akaike weight, evidence ratio) was then used to compare among the demographic models as in Carstens *et al.* (2009).

## Results

### Morphological analyses

Principal components analysis of the mensural data reveals a trend towards differentiation between Buka and Bougainville, although the clusters do broadly overlap (Fig. 3a). Specimens from Buka exhibit slightly lower values for the first principal component (82.1% of variation), which corresponds to shorter limbs relative to body length, but the Buka and Bougainville populations show no substantial differentiation on any other principal components axes. Discriminant function analyses of the mensural data further substantiate this trend of differentiation of the Buka and Bougainville populations based on mensural data: 105 of 114 specimens (92.11%) examined from Bougainville are correctly classified and 37 of 39 specimens (94.87%) examined from Buka are correctly classified based on the mensural data. Meristic data reveal far greater differentiation and diagnosability of the Buka and Bougainville populations. PCA of meristic data (Fig. 3b) shows the populations are primarily differentiated along the first principal component axis (68.9% of variation), with specimens from Bougainville exhibiting larger values, that corresponds both to a larger number of ventral scale rows, and a larger number of subdigital lamellae. DFA based on the meristic data also reveals the diagnosability of the Buka and Bougainville populations: 111 of 114 specimens from Bougainville are correctly

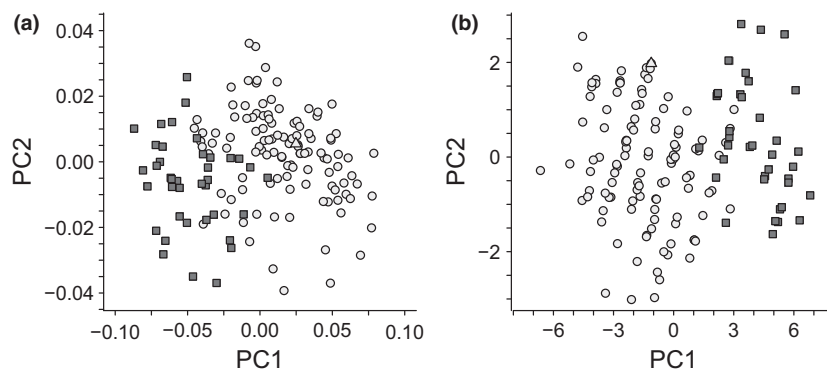
classified (97.37%), and 37 of 39 specimens from Buka are correctly classified (94.87%).

### Next-generation sequencing

The two Ion Torrent PGM 316 chips resulted in a combined 6.45 million reads, or 648.13 MB of sequence data, of which 486.28 MB had an estimated quality score of Q20 or above. After parsing reads by barcoded sample and filtering out low-quality reads, and those with ambiguous bases or errors in the barcode or forward primer sequences, the NGS experiments resulted in an average of 255 748 reads per individual (range 194 980–314 946). A total of 6983 loci were identified using the PRGmatic pipeline, of which 1526 loci were sequenced from all 12 individuals (mean coverage  $62.27 \pm 49.03$  reads per locus per individual), 2608 were sequenced in at least nine individuals (mean coverage  $50.09 \pm 46.61$  reads per locus per individual), and 3342 were sequenced in at least six individuals (mean coverage  $43.21 \pm 44.66$  reads per locus per individual). The data sets including loci sequenced in at least nine individuals were 93.57% complete, whereas those with loci sequenced in at least six individuals were 85.80% complete. Of the 1526 loci sequenced in all 12 individuals, 1002 (65.66%) were variable; 1703 loci of the 2608 sequenced in at least nine individuals (64.99%) were variable; and of 3342 loci sequenced in at least six individuals, 2135 (63.61%) were variable. Among the loci recovered in all 12 individuals, 941 included at least one biallelic SNP and were thus useful for SNAPP and *daði* analyses; 1590 loci recovered in at least nine individuals had at least one biallelic SNP, and 1973 loci recovered in at least six individuals had biallelic SNPs. Among the loci recovered in at least six individuals, the mean number of segregating sites was 0.994 (range: 0–9).

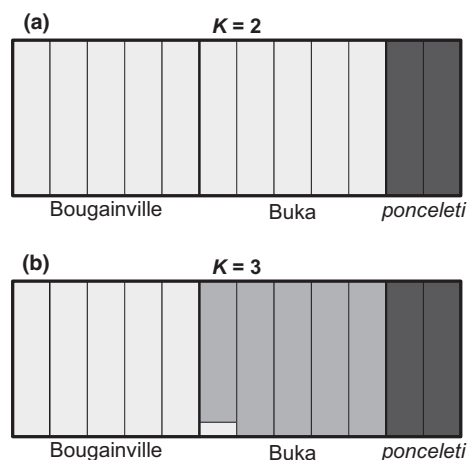
### Genetic clustering analyses

With the complete data set of loci sequenced for all 12 individuals, the  $\ln \Pr(D|K)$  from Structure peaked at



**Fig. 3** Scatterplots of the first two principal components of the morphological data based on (a) mensural characters and (b) meristic characters. Dark grey squares indicate Buka specimens, and light grey circles indicate Bougainville specimens. The holotype is shown as a light grey triangle.

$K = 3$  (Fig. S3, Supporting information).  $\Delta K$  peaked at  $K = 2$  ( $\Delta K = 582.54$ ), with a slightly lower peak at  $K = 3$  ( $\Delta K = 487.60$ ); for all other tested values of  $K$ ,  $\Delta K$  was less than 3.27. With  $K = 2$ , the two clusters corresponded to *T. ponceleti* and *T. pseudoponceleti* with all samples assigned to the appropriate cluster with probability 1.0 (Fig. 4a). With  $K = 3$ , the three clusters corresponded to Bougainville, Buka and *T. ponceleti*, with 11 of the 12 samples assigned to the appropriate cluster with probability of 0.999 or greater (Fig. 4b). The twelfth sample, from Buka, had an admixed assignment, with probability 0.907 to the appropriate cluster, and probability 0.093 to the Bougainville cluster. Structure results were similar when including all loci sequenced for at least nine individuals:  $\ln \Pr(D|K)$  peaked at  $K = 3$ , whereas  $\Delta K$  peaked at  $K = 2$  ( $\Delta K = 474.54$ ), with a slightly lower peak at  $K = 3$  ( $\Delta K = 321.50$ ); for all other tested values of  $K$ ,  $\Delta K$  was less than 6.72. Population assignment probabilities were identical to those obtained with the data set of loci sequenced for all 12 individuals. With the data set of loci sequenced in at least six individuals, both the  $\ln \Pr(D|K)$  and  $\Delta K$  peaked at  $K = 3$  ( $\Delta K = 420.93$ ), with a slightly lower peak for  $\Delta K$  at  $K = 2$  ( $\Delta K = 321.65$ ); for all other tested values of  $K$ ,  $\Delta K$  was less than 3.57. As with the other, more complete data sets with fewer loci, the three clusters corresponded to the Buka, Bougainville and *T. ponceleti* populations; 11 of the 12 individuals were assigned to the appropriate cluster with assignment probability greater than 0.999, with the final sample, an individual from Buka, assigned to the appropriate cluster with probability 0.900, and with probability 0.100 to the Bougainville cluster.



**Fig. 4** Plots of cluster assignments for all samples from Structure analyses of the 1526 loci sequenced in all twelve individuals, produced using *Structure* ver. 1.1 (Rosenberg 2003) with (a)  $K = 2$  and (b)  $K = 3$ .

When examining genetic clustering with the implementation of the Dirichlet process prior to estimate  $K$  in *Structure*, the posterior probability was 1.00 for  $K = 3$ , and 0.00 for all other values of  $K$ , regardless of the data set used (i.e. loci sequenced in all 12 individuals, those sequenced in at least nine individuals or those sequenced in at least six individuals), or the selection of prior distribution on the  $\alpha$  parameter of the Dirichlet process prior. As with the *Structure* analyses, these three clusters corresponded to *T. ponceleti*, and the Buka and Bougainville populations, with all samples assigned appropriately.

#### *Species delimitation using Bayes factors*

Divergence times and effective population sizes were sensitive to the selection of prior on theta: the prior distributions with smaller means resulted in smaller estimated population sizes and shallower divergences (Table S4, Supporting information). However, other parameters, likelihoods and posterior probabilities were not sensitive to prior selection. Similarly, selection of the prior on theta had a negligible impact on the estimation of marginal likelihood. Regardless of the data set used (i.e. complete data set, loci recovered from 9 + individuals or loci recovered from 6+ individuals), the Bayes factors strongly favoured the three species model (i.e. *ponceleti*, Buka, Bougainville; Table 1). This pattern was strongest in the largest data set (i.e. loci recovered from 6+ individuals); however, the smallest data set (including only loci recovered from all 12 individuals), the Bayes factors were over 1480 in favour of the three species model (Table 1). Bayes factors in excess of 10 are typically considered decisive support for a model over the alternative; thus, these results strongly support the treating of Buka and Bougainville as distinct species. Estimated population sizes and divergence times from the full analyses without path sampling are presented in Table S4, Supporting information.

#### *Demographic model selection using genomic data*

$\partial\text{a}\partial\text{i}$  analyses of the AFS strongly support model 9, including population size changes and shift in migration rates subsequent to the divergence between the Buka and Bougainville populations, as the best fit to the data, and encompassing 0.985 of the AIC weight (Table 2). Model 5, including a population size change and a stop of migration subsequent to the divergence between the Buka and Bougainville populations, was the next best fit model; however, this model was a substantially worse fit to the data, including only 0.011 of the AIC weight, and with an evidence ratio of 88.871

**Table 1** Marginal likelihoods and Bayes factors estimates from SNAPP analyses. Positive Bayes factors indicate support for the three species model

Data set	Prior	Marg. Like. 2 sp	Marg. Like. 3 sp	Bayes Factor
Complete	G(2,100)	-7010.6282	-6260.9169	1499.4226
Complete	G(2,1000)	-7006.0223	-6256.3510	1499.3425
Complete	G(2,10000)	-7014.2807	-6270.9888	1486.5837
9 + Ind	G(2,100)	-10745.5990	-9618.7450	2253.7081
9 + Ind	G(2,1000)	-10740.7585	-9613.6024	2254.3123
9 + Ind	G(2,10000)	-10755.0155	-9619.2605	2271.5099
6 + Ind	G(2,100)	-12305.0442	-10847.0262	2916.0360
6 + Ind	G(2,1000)	-12298.1805	-10846.2591	2903.8428
6 + Ind	G(2,10000)	-12313.3828	-10867.6151	2891.5353

**Table 2** Information theoretic statistics of demographic model selection via Akaike information criterion from  $\partial\text{a}\partial\text{i}$  analyses

Model	K	lnL	AIC	$\Delta\text{AIC}$	$w_i$	$E_{\min,i}$
Model 9	13	-449.368	924.736	0.000	0.985	1.000
Model 7	11	-455.855	933.711	8.974	0.011	88.871
Model 5	9	-459.840	937.679	12.943	1.524E-03	646.433
Model 3	8	-460.933	937.866	13.130	1.388E-03	709.695
Model 8	11	-458.769	939.539	14.803	6.014E-04	1638.215
Model 6	11	-459.634	941.269	16.533	2.532E-04	3890.720
Model 2	6	-480.319	972.637	47.901	3.908E-11	2.521E+10
Model 4	9	-480.038	978.076	53.340	2.576E-12	3.825E+11
Model 1	3	-1320.474	2646.947	1722.211	<1E-15	>1E15

relative to model 9. All remaining models had AIC weights of below 0.0016, and evidence ratios in excess of 645, suggesting the other tested models were a far worse fit to the data. Maximum-likelihood parameter estimates suggest the initial divergence between Buka and Bougainville occurred 0.91 (0.57–1.46)  $N_e$  generations ago, while the divergence between these populations and *T. ponceleti* occurred 1.35 (0.75–2.42)  $N_e$  generations ago (Table S5, Supporting information). Based on the estimated relative substitution rate of the NGS loci, these estimates correspond to divergences occurring approximately 0.84 (0.53–1.33) million years before present (Mya) and 1.23 (0.67–2.22) Mya, respectively (Table 3). Parameter estimates further suggest that migration rates were moderate between the Buka and Bougainville populations immediately after divergence (3.943 from Buka to Bougainville, 1.180 from Bougainville to Buka, in units of  $2N_0 m$ , where  $N_0$  is the ancestral effective population size, and  $m$  is the proportion of each population made up of migrants each generation), with a substantial decrease in migration subsequently (0.251 and 0.613, respectively; Table 3). Confidence intervals on migration rates, however, were extremely broad, particularly for migration rates prior to the shift (Table 3, Fig. 5).

The plot of the first two principal components of the summary statistics reveals that the observed summary statistics fall well within the cloud of summary statistics generated by the simulated data (Fig. S2), indicating that the selected prior distributions are appropriate for producing summary statistic vectors comparable to those observed in the empirical data. Pseudo-rejection analyses using simulated summary statistics suggest that, with the exception of model 1 (no divergence between Buka and Bougainville), the summary statistics were generally insufficient to distinguish among models (Table S6, Supporting information). Model 1 was correctly identified with the highest posterior probability in 93% of cases, and in 76% of the replicates, model 1 was correctly identified with moderate support (posterior probability >0.75). Among the remaining models, the correct model had the highest posterior probability in only 37.4% of replicates, and in only six of the 800 total replicates did the correct model receive moderate support. Replicates in which the model with the maximum posterior probability was incorrect typically involved difficulty in distinguishing between the models with no migration (i.e. model 2 or model 3), or among models involving migration (i.e. models 4–9), although models in which migration stopped

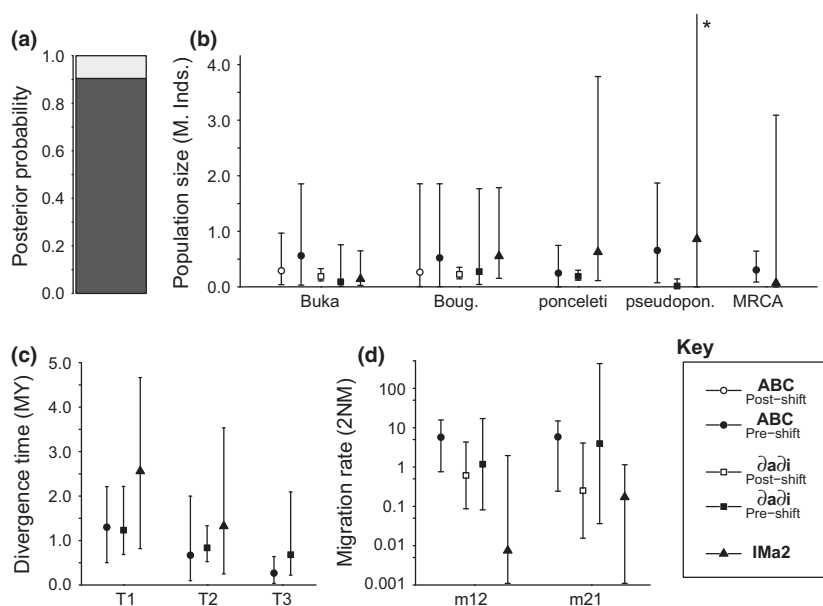
**Table 3** Demographic parameter estimates and 95% HPD intervals (ABC, IMA2) or 95% confidence intervals ( $\hat{\sigma}\hat{\sigma}$ ) from ABC and  $\hat{\sigma}\hat{\sigma}$  analyses of NGS data and IMA2 analysis of Sanger data. Divergence times (T) are given in years, and population size estimates (N) are provided as number of individuals. Relative population size estimates from  $\hat{\sigma}\hat{\sigma}$  analyses were calibrated using the estimated ancestral population size from ABC analyses. Migration rate parameters (m) are provided in units of  $2N_e m$ . See Fig. 2 for details of parameters

	ABC	$\hat{\sigma}\hat{\sigma}$	IMA2
T <sub>1</sub>	1,300,520 (503,200–2,213,182)	1,233,513 (686,373–2,216,661)	2,560,041 (818,693–4,665,252)
T <sub>2</sub>	669,347 (98,132–1,998,519)	836,958 (525,719–1,332,468)	1,323,554 (250,156–3,534,026)
T <sub>3</sub>	267,208 (37,074–639,687)	681,521 (221,895–2,093,630)	—
N <sub>1b</sub>	291,434 (40,003–967,499)	188,544 (107,988–329,150)	553,918 (154,317–1,785,206)
N <sub>1a</sub>	561,837 (32,279–1,854,005)	92,883 (11,382–758,172)	
N <sub>2b</sub>	265,215 (5,599–1,855,487)	226,625 (145,519–352,890)	142,297 (26,640–649,107)
N <sub>2a</sub>	523,069 (5,599–1,855,487)	276,026 (43,146–1,766,219)	
N <sub>3</sub>	248,499 (1,169–746,915)	189,795 (120,376–299,216)	627,015 (113,707–3,784,832)
N <sub>12</sub>	656,649 (75,675–1,868,975)	16,233 (1,831–143,292)	860,928 (0–30,489,829)
N <sub>a</sub>	305,136 (88,026–643,408)	—	70,174 (0–3,090,242)
m <sub>12b</sub>	0	0.6126 (0.0869–4.3169)	0.007496 (0.00–1.957)
m <sub>12a</sub>	11.4474 (1.5165–31.4366)	1.1797 (0.0816–17.0565)	
m <sub>21b</sub>	0	0.2512 (0.0155–4.0678)	0.1717 (0.00–1.142)
m <sub>21a</sub>	11.6862 (0.4857–29.6822)	3.9435 (0.0364–426.8419)	

subsequent to the divergence (model 5 or model 7) were difficult to distinguish from either isolation models or isolation with migration models (Table S6, Supporting information). However, when the model was not identified correctly, all models generally received poor support, and in only a single replicate did an incorrect model have posterior support greater than 0.75, thus suggesting that when the posterior probability strongly supported a single model, this was typically the correct model. Parameter estimation using pseudo-observed data performed better, with 97.8% of the replicates including the true value within the 95% HPD intervals, and 68.1% of replicates including the true value within a single standard deviation of the geometric mean (Table S7, Supporting information). However, 95% HPD intervals for most parameters included the majority of the prior distribution, although with smaller variance.

ABC analyses of the empirical data resulted in a high posterior probability for model 7 (PP = 0.9043), which

includes a shift in population size and cessation of migration subsequent to the divergence of Buka and Bougainville. Model 9, involving a shift in population size and in migration rate subsequent to the divergence of Buka and Bougainville, also received some limited posterior support (PP = 0.0957); the remaining models received little support, with posterior probabilities less than 0.0001. Similar to the pseudo-observed data, parameter estimation resulted in broad 95% HPD intervals and posterior distributions similar in width to, but more tightly clustered than, prior distributions (Fig. 5, Fig. S4, Supporting information). ABC analyses estimate divergence times similar to those estimated in the  $\hat{\sigma}\hat{\sigma}$  analyses: 0.73 (0.11–2.18)  $N_e$  generations ago between Buka and Bougainville and 1.42 (0.55–2.42)  $N_e$  generations ago between these populations and *T. poncetii* (Table S5, Supporting information). Using the mitochondrial calibration rate, these correspond to 0.67 (0.098–2.00) Mya and 1.30 (0.50–2.21) Mya, respectively (Table 3). Estimates of the migration rate parameters,



**Fig. 5** (a) Posterior probabilities of demographic models estimated via ABC. Dark grey indicates model 7, and light grey indicates model 9. All other models had posterior probabilities  $<0.0001$ . (b) Estimated effective population sizes, in millions of individuals. (c) Estimated timing of demographic events in millions of years: T1 is the divergence of *ponceleti* and Buka + Bougainville, T2 is the divergence of Buka and Bougainville, and T3 is the timing of the shift in population size and migration rates. (d) Estimated migration rates, in units of  $2N_0m_{ij}$ , from Bougainville to Buka (m12) and Buka to Bougainville (m21). For sections b–d, error bars indicate the limits of the 95% confidence intervals ( $\hat{\sigma}a\hat{\sigma}i$  analyses) or the 95% HPD intervals (ABC, IMA2 analyses). \*The upper limit on effective population size for the common ancestor of Buka + Bougainville (pseudopon.) based on IMA2 analyses was 30 489 829 individuals.

**Table 4** Information theoretic statistics for demographic model selection via Akaike information criterion from IMA2 analyses of Sanger data

Model	$K$	$\log(P)$	AIC	$\Delta AIC$	$w_i$	$E_{\min,i}$
No Migration	7	-1.271	16.542	0.000	1.000	1.000
Migration Buka to Boug. Only	8	-1.271	18.542	2.000	0.368	2.718
Symmetrical Migration	8	-1.799	19.598	3.056	0.217	4.609
Migration Boug. to Buka Only	8	-1.914	19.828	3.286	0.193	5.171
Full Model	9	-1.271	20.542	4.000	0.135	7.389

however, were difficult to assess via ABC analyses and yielded posterior distributions similar to the prior distributions (Fig. S4, Supporting information).

#### Demographic modelling using Sanger data

The maximized posterior density function estimates ( $-\log(P)$ ) from IMA2 were highly similar among the five models tested (Table 4); thus, the AIC supported the least parameter-rich model of pure isolation with no migration between Buka and Bougainville. However, the largest difference in AIC score among the models was 4.0, and the relative model probabilities (AIC weights) were all in excess of 0.13, indicated limited

support for the simplest model over the alternatives. Demographic parameter estimates from IMA2 analyses were similar to those based on the NGS data, through generally with much broader confidence intervals, particularly for divergence times, and population sizes of *T. ponceleti* and the ancestral populations (Fig. 5, Table 3). The divergence between the Buka and Bougainville populations was estimated as 0.51 (0.096–1.36)  $N_e$  generations, or 1.32 (0.25–3.53) Mya, while the divergence between *T. ponceleti* and these populations was 0.99 (0.32–1.80)  $N_e$  generations, or 2.56 (0.82–4.67) Mya (Table 3, Table S5, Supporting information). Based on the full isolation with migration model, IMA2 analyses of the Sanger data estimate migration rates of 0.0075

from Buka to Bougainville and 0.172 from Bougainville to Buka, in units of  $2N_e m$  (Table 3).

## Discussion

Although species are the most basic unit of biological classification and as such are fundamental to all biological research, there remains much contention over species concepts and what criteria should be used in identifying species. This controversy likely stems at least in part from the variability and complexity in the underlying processes that lead to speciation (Coyne & Orr 2004). The general lineage species concept acknowledges this variation and leverages it to a more general definition of species as independently evolving population lineages, allowing researchers to apply the specific criterion or criteria most applicable to their system. While in many systems, species boundaries may be robustly inferred with less extensive data and analyses than we present, through combining multiple data sets and analyses, we are able to validate the speciation of the Buka and Bougainville populations of *Tribolonotus pseudoponzeleti* under a variety of criteria and thus provide a far more comprehensive assessment of the status of these populations. Our previous analyses (Austin *et al.* 2010) found the samples from Buka and Bougainville to be reciprocally monophyletic on the basis of 2252 bp of mitochondrial and nuclear sequence data (2 mitochondrial and 3 nuclear gene regions), but did not specifically test these populations for morphological differences, examine the demographic history of these populations, or specifically test if this divergence is indicative of speciation between these populations. The morphological data we present here indicate the presence of diagnostic morphological characters distinguishing the two populations (body size, number of ventral scale rows, number of subdigital lamellae). DFAs were accurate in distinguishing the Buka and Bougainville populations on the basis of the meristic data, and, to a lesser extent, on the basis of mensural data, further supporting the morphological diagnosability of the populations. PCAs based on meristic data also show the two populations generally form distinct clusters; again to a lesser extent, the populations largely cluster in principal components space on the basis of mensural data. Genetic clustering analyses further support the distinctness of the Buka and Bougainville populations. While  $\Delta K$  peaks at  $K = 2$  for the smaller (but more complete) data sets, with the two clusters corresponding to *T. ponzeleti* and *T. pseudoponzeleti*, both show a slightly lower secondary peak at  $K = 3$ , with the clusters corresponding to *T. ponzeleti*, Buka, and Bougainville. In cases of hierarchical population structure,  $\Delta K$  detects higher levels of clustering first (Evanno *et al.* 2005); thus, this

likely reflects the hierarchical nature of the data (i.e. the Buka and Bougainville populations are far more closely related to each other than either is to *T. ponzeleti*), rather than evidence for a single *T. pseudoponzeleti* cluster. Further, with the data set consisting of loci sequenced in at least six samples,  $\Delta K$  peaks at  $K = 3$ . Structurama analyses always resulted in a posterior probability of 1.00 for  $K = 3$ , regardless of the data set used or the prior distribution. Coalescent estimates of the species tree in SNAPP strongly prefer the three species model over two species, regardless of the selection of prior on theta and the completeness of the data set. Finally, demographic analyses strongly reject a model of no divergence between Buka and Bougainville and suggest that, while gene flow has likely occurred between these populations following divergence, this migration has subsequently decreased substantially ( $\partial a \partial i$ ), or ceased completely (ABC). Further, other scincid lizards with comparable differences in body size (*Plestiodon skiltonianus* group) have been shown to be reproductively isolated as a direct result of body size differences preventing successful copulation (Richmond & Jockusch 2007). While the copulatory posture of *Tribolonotus* is unknown, it is plausible a similar mechanism acts as a partial prezygotic barrier to reproduction between the two island populations. These data thus corroborate each other and provide comprehensive evidence that the Buka and Bougainville populations represent distinct, evolutionarily independent species. These data further support the species-level differentiation of these populations under a variety of other species concepts, including the genealogical species concept (Baum & Shaw 1995), the diagnostic or phylogenetic species concept (Cracraft 1989), the phenetic species concept (Sokal & Crovello 1970), the genotypic or genomic species concept (Mallet 1995, 2001; Bradley & Baker 2001; Baker & Bradley 2006) and, arguably, the biological species concept (Mayr 1942, 1970). While the description of this new species of *Tribolonotus* from Buka Island is beyond the scope of this study, from a biodiversity and conservation standpoint it is critically important that the new species be described. Thus, we are currently preparing an additional manuscript to describe this new species.

This divergence between the Buka and Bougainville populations is particularly remarkable given their geographic context. The Buka Passage that separates the two islands is only approximately 300 m in width, and the islands were repeatedly merged during Pleistocene glacial cycles, including as recently as 10 000–20 000 years ago (Chappel & Shackleton 1986; Mayr & Diamond 2001). The few studies have examined phylogeographic patterns of terrestrial species in the Solomon Archipelago and have found varied levels of divergence among islands, with some taxa showing

limited divergence and others showing deep divergence among islands (Filardi & Smith 2005; Pulvers & Colgan 2007; Smith & Filardi 2007; Hagen *et al.* 2012). However, these studies have not included samples from both Buka and Bougainville and have focused on species that likely have relatively high dispersal capabilities (e.g. birds, bats, or large, arboreal lizards). In contrast, *Tribolonotus* are semifossorial, secretive species, rarely found outside cover material (e.g. decaying logs), and likely susceptible to desiccation (Greer & Parker 1968; McCoy 2006). Thus, although no estimates of dispersal distances in *Tribolonotus* are available, it is reasonable to expect these species are extremely dispersal limited. Many species in the Solomon Archipelago are known from multiple island groups, including islands far more isolated than Buka and Bougainville, and with no history of connectivity during Pleistocene glacial cycles (Mayr & Diamond 2001; McCoy 2006; Menzies 2006). These results therefore suggest that some such species, particularly those with limited vagility, may instead represent complexes of multiple distinct taxa and thus vastly underestimate biodiversity of the region.

Demographic analyses reveal further details on the underlying mechanisms responsible for the speciation between Buka and Bougainville. While ABC pseudo-observed analyses suggest model selection is difficult, the two models receiving posterior support were also identified as best fit models by  $\delta a \delta i$  analyses, lending further credibility to this model. This model is also reasonable given the geological history of repeated glaciations, connecting the two islands and likely enabling moderate levels of gene flow, and interglacial cycles, isolating the islands and likely limiting or preventing migration. The divergence between Buka and Bougainville was also estimated to have occurred during the mid-Pleistocene (669 kya, 837 kya or 1.32 Mya based on ABC,  $\delta a \delta i$  or IMA2 analyses, respectively; Table 3), supporting the hypothesis that early glacial cycles may have played an important role in driving the divergence between these populations.

Despite the strong evidence for the inclusion of migration parameters based on the genomic data, IMA2 analyses of the Sanger data fail to reject the simplest model of no migration between Buka and Bougainville, and all tested models received significant support (minimum relative model probability = 0.13). However, previous power analyses using a similar number of loci and similar demographic parameters found that, for data simulated under a full isolation with migration model (i.e. including migration), pure isolation models could not be rejected in approximately 40% of the replicates (Carstens *et al.* 2009). Therefore, it is plausible that the differences in the importance of migration between the analyses based on Sanger data and those based on

the genomic data may reflect the limited power to distinguish between pure isolation and isolation with migration models based on a small number of loci, rather than differing signals between the data sets.

Parameter estimates were generally similar across demographic analyses, and no significant differences were detected between  $\delta a \delta i$  and ABC analyses of the genomic data, or between these analyses and the IMA2 analysis of the Sanger data (Fig. 5, Table 3). Confidence intervals on parameter estimates, however, were generally much broader in analysis based on the Sanger data relative to the genomic analyses, particularly for the divergence time estimates, the population size of *T. ponceleti* and of the ancestral population sizes (Fig. 5, Table 3). The much greater breadth of the confidence interval on the population size of *T. ponceleti* (for which our sample size was smallest) based on the Sanger data relative to the genomic analyses is particularly intriguing and suggests that large numbers of independent loci may mitigate the limitations of small sample sizes. Combined, these data further support the results of previous studies (e.g. (Carling & Brumfield 2007; McCormack *et al.* 2009; Smith *et al.* 2014) that suggest that the number of independent loci is the most important aspect of sampling strategy for coalescent analyses.

The proposed scenario, as estimated via the demographic analyses, posits no necessity for ecological divergence between these species, although it does not preclude this possibility. Indeed, differences in limb lengths, digit lengths and body size have all been shown to correlate with ecological divergence in lizards (Melville & Swain 2000; Kohlsdorf *et al.* 2001; Goodman 2007). However, the ecology of *Tribolonotus* is poorly known, and beyond the morphological differences described here, no data are available to test whether ecological divergence may accompany the speciation of the Buka and Bougainville populations. More detailed studies aimed specifically at examining the ecology of these species are necessary to assess whether the divergence between Buka and Bougainville occurred solely as a result of allopatric divergence, or whether ecological divergence plays an important role along with allopatry in driving and maintaining speciation in this group.

The shallow divergence between the Bougainville and Buka populations illustrates the power of NGS data to elucidate evolutionary processes in recently diverged species groups. In spite of this power, it is through combining this genomic data with morphological data that we are able to assess the validity of the Buka and Bougainville populations under a variety of species concepts and thus provide robust evidence of the status of these populations as distinct species. While many systems do not require a genomic data set of the scale collected in this study for the accurate and robust

validation of putative species, this study provides a framework for how next-generation sequencing data can be combined with morphological data and leveraged for species validation, and how further analyses of genomic data can facilitate elucidating more detailed demographic processes associated with speciation.

## Acknowledgements

We thank Cathy Newman, Jeff Weinell and three anonymous reviewers for providing useful comments on earlier drafts of this manuscript, as well as Adam Leaché and Remco Bouckart for assistance with SNAPP analyses, Michael Harvey for assistance with *â* analyses and Sara Hird for assistance with PRGmatic. Steve Richards provided tissues and voucher specimens of *Tribolonotus poncelleti*, and Ross Sadler, Kathleen Imada, and Allen Allison provided specimen loans. Jonathan Losos, José Rosado, Tsuyoshi Takahashi, Jonathan Woodward, David Kizirian and Christopher Raxworthy were all helpful in facilitating museum collections visits. This research was funded by a Harvard University Museum of Comparative Zoology Ernst Mayr Travel Grant in Animal Systematics to ENR and National Science Foundation grant DEB 1146033 to CCA.

## References

- Arbogast BS, Kenagy GJ (2008) Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, **28**, 819–825.
- Austin CC, Rittmeyer EN, Richards SJ, Zug GR (2010) Phylogeny, historical biogeography and body size evolution in Pacific Island Crocodile skinks, *Tribolonotus* (Squamata; Scincidae). *Molecular Phylogenetics and Evolution*, **57**, 227–236.
- Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, **87**, 643–662.
- Bauer AM, Jackman T, Smith SA, Saidler RA, Austin CC (2004) Note: vocalization in *Nannoscincus gracilis* (New Caledonian Gracile Dwarf Skink). *Herpetological Review*, **35**, 268–269.
- Baum DA (2009) Species as ranked taxa. *Systematic Biology*, **58**, 74–86.
- Baum DA, Shaw KL (1995) Genealogical Perspectives on the Species Problem. In: *Experimental Approaches to Plant Systematics – Monographs in Botany from the Missouri Botanical Garden 53* (eds Hoch PC, Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, Missouri.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.
- Bouckaert R, Heled J, Kühnert D *et al.* (2013) Beast2: A software platform for Bayesian evolutionary analysis.
- Bradley RD, Baker RJ (2001) A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy*, **82**, 960–973.
- Brown RP, Pestano J (1998) Phylogeography of skinks (*Chalcides*) in the Canary Islands inferred from mitochondrial DNA sequences. *Molecular Ecology*, **7**, 1183–1191.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.
- Burbrink FT (2001) Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). *Herpetological Monographs*, **15**, 1–53.
- Capriglione T, Olmo E, Odierna B, Improta B, Morescalchi A (1987) Cytofluorometric DNA base determination in vertebrate species with different genome sizes. *Basic and Applied Histochemistry*, **31**, 119–126.
- Carling MD, Brumfield RT (2007) Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLoS ONE*, **2**, e160.
- Carranza S, Arnold EN (2003) Investigating the origin of transoceanic distributions: Mtdna shows *Mabuya* lizards (Reptilia, Scincidae) crossed the Atlantic twice. *Systematics and Biodiversity*, **1**, 275–282.
- Carstens BC, Stoute HN, Reid NM (2009) An information-theoretic approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.
- Chappel J, Shackleton NJ (1986) Oxygen isotopes and sea level. *Nature*, **324**, 137–140.
- Chappell J, Omura A, Esat T *et al.* (1996) Reconciliation of late Quaternary sea levels derived from coral terraces at Huon Peninsula with deep sea oxygen isotope records. *Earth and Planetary Science Letters*, **141**, 227–236.
- Cogger HG (1972) A new scincid lizard of the genus *Tribolonotus* from Manus Island, New Guinea. *Zoologische Mededelingen (Leiden)*, **47**, 201–210.
- Cole JR, Wang Q, Cardenas E *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research*, **37**, D141–D145.
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates Inc, Sunderland, Massachusetts.
- Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its Consequences* (eds Otte D, Endler JA), pp. 28–59. Sinauer Associates Inc, Sunderland, Massachusetts.
- Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.
- Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.
- De Queiroz K (2005) A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences*, **56**, 196–215.
- De Queiroz K (2007) Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.
- De Smet WHO (1981) The nuclear Feulgen-DNA content of the vertebrates (especially reptiles), as measured by fluorescence cytophotometry, with notes on the cell and chromosome size. *Acta Zoologica et Pathologica Antverpiensia*, **76**, 119–167.
- Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–1797.



- 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.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Filardi CE, Smith CE (2005) Molecular phylogenetics of monarch flycatchers (genus *Monarcha*) with emphasis on Solomon Island endemics. *Molecular Phylogenetics and Evolution*, **37**, 776–788.
- Goodman BA (2007) Divergent morphologies, performance, and escape behaviour in two tropical rock-using lizards (Reptilia: Scincidae). *Biological Journal of the Linnean Society*, **91**, 85–98.
- Greer AE, Parker F (1968) A new species of *Tribolonotus* (Lacertilia: Scincidae) from Bougainville and Buka, Solomon Islands, with comments on the biology of the genus. *Breviora*, **291**, 1–23.
- Gregory TR (2014) Animal Genome Size Database. <http://www.genomesize.com>.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics*, **5**, e1000695.
- Hagen IJ, Donnellan SC, Bull CM (2012) Phylogeography of the prehensile-tailed skink *Corucia zebrata* on the Solomon Archipelago. *Ecology and Evolution*, **2**, 1220–1234.
- Hartdegen RW, Russell MJ, Young B, Reams RD (2001) Vocalization of the Crocodile Skink, *Tribolonotus gracilis* (De Rooy, 1909), and evidence of parental care. *Contemporary Herpetology*, **2001**, 1–9.
- Hausdorf B (2011) Progress toward a general species concept. *Evolution*, **65**, 923–931.
- Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.
- Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.
- Hird SM, Brumfield RT, Carstens BC (2011) PRGmatic: an efficient pipeline for collating genome-enriched second-generation sequencing data using a “provisional-reference genome”. *Molecular Ecology Resources*, **11**, 743–748.
- Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. *Genetics*, **175**, 1787–1802.
- Huelsenbeck JP, Andolfatto P, Huelsenbeck ET (2011) Structurama: bayesian inference of population structure. *Evolutionary Bioinformatics*, **7**, 55–59.
- Jackson ND, Austin CC (2010) The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution*, **64**, 409–428.
- Jackson ND, Austin CC (2012) Inferring the evolutionary history of divergence despite gene flow in a lizard species, *Scincella lateralis* (Scincidae), composed of cryptic lineages. *Biological Journal of the Linnean Society*, **107**, 192–209.
- Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Kohlsdorf T, Garland T Jr, Navas CA (2001) Limb and tail lengths in relation to substrate usage in *Tropidurus* Lizards. *Journal of Morphology*, **248**, 151–164.
- Lance SL, Hagen C, Glenn TC *et al.* (2008) Fifteen polymorphic microsatellite loci from Jamaican streamertail hummingbirds (*Trochilus*). *Conservation Genetics*, **10**, 1195–1198.
- Lartillot N, Philippe H (2006) Computing Bayes factors using thermodynamic integration. *Systematic Biology*, **55**, 195–207.
- Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534–542.
- MacCulloch RD, Upton DE, Murphy RW (1996) Trends in nuclear DNA content among amphibians and reptiles. *Comparative Biochemistry and Physiology*, **113B**, 601–605.
- Malhotra A, Thorpe RS (2000) The dynamics of natural selection and vicariance in the Dominican Anole: patterns of within-island molecular and morphological divergence. *Evolution*, **54**, 245–258.
- Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.
- Mallet J (2001) Species, Concepts of. In: *Encyclopedia of Biodiversity* (ed. Levin S), pp. 427–440. Academic Press, Elsevier, Oxford, UK.
- Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York City, New York.
- Mayr E (1970) *Populations, Species, and Evolution*. Harvard University Press, Cambridge, Massachusetts.
- Mayr E, Diamond JM (2001) *The Birds of Northern Melanesia: Speciation, Ecology, and Biogeography*. Oxford University Press, New York, USA.
- McCormack JE, Huang H, Knowles LL (2009) Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology*, **58**, 501–508.
- McCormack JE, Maley JM, Hird SM *et al.* (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution*, **62**, 397–406.
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.
- McCoy M (2006) *Reptiles of the Solomon Islands*. Pensoft Publishing, Sofia-Moscow.
- Melville J, Swain R (2000) Evolutionary relationships between morphology, performance and habitat openness in the lizard genus *Niveoscincus* (Scincidae: Lygosominae). *Biological Journal of the Linnean Society*, **70**, 667–683.
- Menzies JI (2006) *Frogs of New Guinea and the Solomon Islands*. Pensoft Publishers, Sofia, Bulgaria.
- Metzker ML (2010) Sequencing technologies – the next generation. *Nature Reviews. Genetics*, **11**, 31–46.
- Naomi S-I (2011) On the integrated frameworks of species concepts: Mayden’s hierarchy of species concepts and de Queiroz’s unified concept of species. *Journal of Zoological Systematics and Evolutionary Research*, **49**, 177–184.
- O’Meara BC (2010) New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, **59**, 59–73.

- Pavlidis P, Laurent S, Stephan W (2010) msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Molecular ecology resources*, **10**, 723–727.
- Plummer M, Best N, Cowles K, Vines K (2005) coda: Output Analysis and Diagnostics for MCMC.
- Poulakakis N, Lymberakis P, Valakos E *et al.* (2005) Phylogeography of Balkan wall lizard (*Podarcis taurica*) and its relatives inferred from mitochondrial DNA sequences. *Molecular Ecology*, **14**, 2433–2443.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Pulvers JN, Colgan DJ (2007) Molecular phylogeography of the fruit bat genus *Melonycteris* in northern Melanesia. *Journal of Biogeography*, **34**, 713–723.
- Pyron RA, Burbrink FT (2009) Systematics of the Common Kingsnake (*Lampropeltis getula*; Serpentes: Colubridae) and the burden of heritage in taxonomy. *Zootaxa*, **2241**, 22–32.
- Revelle W (2012) psych: Procedures for Personality and Psychological Research.
- Rheindt FE, Székely T, Edwards SV *et al.* (2011) Conflict between genetic and phenotypic differentiation: the evolutionary history of a “lost and rediscovered” shorebird. *PLoS ONE*, **6**, e26995.
- Rheindt FE, Fujita MK, Wilton PR, Edwards SV (2014) Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): population genetic and phylogenetic inferences from Genome-Wide SNPs. *Systematic Biology*, **63**, 134–152.
- Richmond JQ, Jockusch EL (2007) Body size evolution simultaneously creates and collapses species boundaries in a clade of scincid lizards. *Proceedings of the Royal Society B*, **274**, 1701–1708.
- Rittmeyer EN, Austin CC (2012) The effects of sampling on delimiting species from multi-locus sequence data. *Molecular phylogenetics and evolution*, **65**, 451–463.
- Rosenberg NA (2003) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
- Shackleton NJ (1987) Oxygen isotopes, ice volume and sea level. *Quaternary Science Reviews*, **6**, 183–190.
- Smith CE, Filardi CE (2007) Patterns of molecular and morphological variation in some Solomon Island land birds. *The Auk*, **124**, 479.
- Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic biology*, **63**, 83–95.
- Sokal RR, Crovello TJ (1970) The biological species concept: a critical evaluation. *The American Naturalist*, **104**, 127–153.
- Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews. Genetics*, **14**, 404–414.
- Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American journal of human genetics*, **73**, 1162–1169.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York city, New York.
- Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.
- Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
- Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264–9269.
- Zweifel RG (1966) A new lizard of the genus *Tribolonotus* (Scincidae) from New Britain. *American Museum Novitates*, **2264**, 1–12.

---

E.N.R. designed the study, collected morphological and molecular data, analysed and interpreted the data and wrote the article. C.C.A. conducted fieldwork, funded the laboratory work and provided substantial intellectual contributions that vastly improved the study and the final manuscript. Both authors read and approved the final manuscript.

---

### Data accessibility

Raw next-generation sequence data processed locus and alleles calls from PRGmatic, and morphological data have been deposited in Dryad repository (doi:10.5061/dryad.87550). Scripts for preparing structure and structurama input files from PRGmatic locus and allele calls (seq2-struct), for calling biallelic SNPs for SNAPP and  $\hat{\alpha}$  analyses (biSNPcaller), and for conducting simulations for demographic analyses via ABC, are also available via the Dryad repository (doi:10.5061/dryad.87550). Finally, example input files and R scripts have been deposited in the Dryad repository (doi:10.5061/dryad.87550).

### Supporting information

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

**Fig. S1** Schematic of the nine demographic models tested.

**Fig. S2** Plot of first two principal components (PC) from the summary statistics of 900 simulated datasets (100 for each of the three models), and the observed summary statistics.

**Fig. S3** (A) Plot of  $\Delta K$  for dataset of loci recovered from all twelve individuals. (B) Plot of  $\ln \Pr(D|K)$  for dataset of loci recovered from all twelve individuals. Blue circles indicate the mean value for each  $K$ , and grey lines show the standard deviation. (C) Plot of  $\Delta K$  for dataset of loci sequenced in at least nine individuals. (D) Plot of  $\ln \Pr(D|K)$  for dataset of loci sequenced in at least nine individuals. Blue circles indicate the

mean value for each  $K$ , and grey lines show the standard deviation. (E) Plot of  $\Delta K$  for dataset of loci sequenced in at least six individuals. F. Plot of  $\ln \Pr(D|K)$  for dataset of loci sequenced in at least six individuals. Blue circles indicate the mean value for each  $K$ , and grey lines show the standard deviation.

**Fig. S4** Density plots of the posterior distributions of all parameters from the ABC analyses.

**Table S1** Appendix. Specimens included in molecular analyses, and in morphological analyses. The holotype of *Tribolonotus pseudoponceleti* (collected in Kunua, Bougainville Island) is shown in bold.

**Table S2** Sequences of adapters and primers used in next-generation library preparation.

**Table S3** Genbank accession numbers for the previously collected Sanger sequence data used in IMA2 analyses.

**Table S4** Estimated divergence times (in coalescent units) and thetas from SNAPP analyses under different priors on theta for each dataset of varying completeness.

**Table S5** Uncalibrated demographic parameter estimates and 95% HPD intervals (ABC, IMA2) or 95% confidence intervals ( $\hat{\theta}$ ) from ABC and  $\hat{\theta}$  analyses of NGS data and IMA2 analysis of Sanger data.

**Table S6** Accuracy of approximate Bayesian computation model prediction, based on performing rejection steps on 100 randomly selected suites of summary statistics under each of the models.

**Table S7** Accuracy of parameter estimation from approximate Bayesian computation, based on performing rejection steps on 100 simulated suites of summary statistics.