ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil

Marcelo Gehara[1] [ID] | Adrian A. Garda[2] | Fernanda P. Werneck[3] |
Eliana F. Oliveira[4,5] | Emanuel M. da Fonseca[4] | Felipe Camurugi[6] |
Felipe de M. Magalhães[6] | Flávia M. Lanna[4] | Jack W. Sites Jr[7] | Ricardo Marques[6] |
Ricardo Silveira-Filho[6] | Vinícius A. São Pedro[4,8] | Guarino R. Colli[9] | Gabriel
C. Costa[10] | Frank T. Burbrink[1]

[1]Department of Herpetology, American Museum of Natural History, New York, NY, USA

[2]Departamento de Botânica e Zoologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Lagoa Nova, Natal, RN, Brazil

[3]Coordenação de Biodiversidade, Programa de Coleções Científicas Biológicas, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, AM, Brazil

[4]Programa de Pós-Graduação em Ecologia, Universidade Federal do Rio Grande do Norte, Lagoa Nova, Natal, RN, Brazil

[5]Centro de Ciências Biológicas e da Saúde, Laboratório de Zoologia, Cidade Universitária, Universidade Federal do Mato Grosso do Sul, Campo Grand, MS, Brazil

[6]Programa de Pós-Graduação em Ciências Biológicas (Zoologia), Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, PB, Brazil

[7]Department of Biology and Bean Life Science Museum, Brigham Young University, Provo, UT, USA

[8]Centro de Ciências da Natureza, Universidade Federal de São Carlos, Buri, SP, Brazil

[9]Departamento de Zoologia, Universidade de Brasília, Brasília, Brazil

[10]Department of Biology, Auburn University at Montgomery, Montgomery, AL, USA

Correspondence
Marcelo Gehara, Department of Herpetology, American Museum of Natural History, New York, NY, USA.
Email: marcelo.gehara@gmail.com

Abstract

Many studies propose that Quaternary climatic cycles contracted and/or expanded the ranges of species and biomes. Strong expansion–contraction dynamics of biomes presume concerted demographic changes of associated fauna. The analysis of temporal concordance of demographic changes can be used to test the influence of Quaternary climate on diversification processes. Hierarchical approximate Bayesian computation (hABC) is a powerful and flexible approach that models genetic data from multiple species, and can be used to estimate the temporal concordance of demographic processes. Using available single-locus data, we can now perform large-scale analyses, both in terms of number of species and geographic scope. Here, we first compared the power of four alternative hABC models for a collection of single-locus data. We found that the model incorporating an a priori hypothesis about the timing of simultaneous demographic change had the best performance. Second, we applied the hABC models to a data set of seven squamate and four amphibian species occurring in the Seasonally Dry Tropical Forests (Caatinga) in northeastern Brazil, which, according to paleoclimatic evidence, experienced an increase in aridity during the Pleistocene. If this increase was important for the diversification of associated

xeric-adapted species, simultaneous population expansions should be evident at the community level. We found a strong signal of synchronous population expansion in the Late Pleistocene, supporting the increase of the Caatinga during this time. This expansion likely enhanced the formation of communities adapted to high aridity and seasonality and caused regional extirpation of taxa adapted to wet forest.

1 | INTRODUCTION

The extraordinary species diversity in tropical biomes has been explained by higher speciation rates, lower extinction rates, larger area, higher stability, older age and higher rates of dispersal (Brown, 2014). While most of these processes operate over exceedingly old times, several studies suggest that Quaternary climatic changes have played a major role in generating current diversity patterns (Graham, Moritz, & Williams, 2006; Hewitt, 2000). Quaternary climatic changes likely caused contraction and expansion of species and biome ranges. In regions that experienced drastic changes, such as the Eastern Nearctic, climate generated profound changes in population sizes (Burbrink et al., 2016). Some studies show that regions that experienced less climatic variation have high genetic diversity due to lower extinction rates and maintenance of larger long-term population sizes (e.g., Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009). If climatic changes had a major influence on diversity, such as causing contraction and expansion of entire biomes, then synchronous demographic changes are expected across codistributed populations (i.e., community). Thus, investigating the demographic history of communities can determine how much influence past climatic change had on biogeographical processes.

In the last decade, data sets and methods that model historical demographic change have moved from single-locus to genomic scales (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Xue & Hickerson, 2015), while large collections of single-locus data continued to accumulate in online repositories (Garrick et al., 2015). Single-locus data, although not very informative for single-species inference (Degnan & Rosenberg, 2009), may provide important insights about evolution and biogeography if combined into a single comparative framework (Hickerson, Stahl, & Lessios, 2006). Thus, the availability of "community" single-locus data provides an opportunity for large-scale analyses, both in terms of number of species and geographic scope. In this context, methods of hierarchical approximate Bayesian computation (hABC) represent a powerful and flexible approach to jointly analyse multispecies genetic data from large geographic areas (Hickerson, Stahl, & Takebayashi, 2007). Recently, Chan, Schanzenbach, and Hickerson (2014) proposed an hABC method to analyse the historical demography of a community using single-locus data. Their hABC approach simulates a model of demographic change across n-populations, drawing from uniform distributions: the hyper-parameter zeta (ζ), which represents the proportion of simultaneous

demographic change in a community, and the time of simultaneous demographic change (Ts) (Figure 1a). When ζ equals one, all population size changes occur at the same time. Values lower than one represent different levels of demographic synchronism which decreases towards zero. To simulate idiosyncratic expansion ($\zeta < 1$), population-specific expansion times (t_n) are sampled from the same uniform distribution of Ts (Figure 1a). By performing an ABC rejection/regression on observed and simulated hypersummary statistics, the method estimates: the proportion of simultaneous demographic change (ζ), the time of simultaneous change (Ts), the mean time of all demographic changes [$E(t)$] and dispersion across all times (DI; see Chan et al., 2014 and Methods for details).

One important aspect of the Chan et al. (2014) method is that no minimum temporal difference is considered when modelling idiosyncratic demographic changes. Two events would be considered idiosyncratic regardless of whether they occurred a hundred years apart or 100,000 years apart. However, it may be incorrect to assume these two examples as equally idiosyncratic. Fine temporal differences are likely hard to assess with single-locus data and could probably be ignored. From a historical biogeography perspective, these differences may be also irrelevant. It might be more appropriate to model an a priori hypothesis for the timing of synchronous demographic changes and to constrain the temporal difference between idiosyncratic events. Hierarchical models that implement these aspects might fit the data better and increase the variance among summary statistics generated under different zeta values, improving the accuracy of posterior estimates.

Here, we implement alternative parameterizations of the demographic history of a community. We enforce a time difference between Ts and t_n to model idiosyncratic changes, and include the possibility of defining an a priori hypothesis for the timing of synchronous demographic changes. We evaluated precision and accuracy of our implementation and the method of Chan et al. (2014) using pseudo-observed data (POD). We further used these new hABC models to conduct a comparative demographic analysis for a community of 11 species of amphibians and squamates from northeastern Brazil, where demographic histories of communities are relatively poorly known throughout the Pleistocene.

The distribution of biomes in northeastern Brazil is consistent with alternating contraction/expansion cycles of wet forest and dry formations. The region is primarily covered by a large block of Seasonally Dry Tropical Forests (SDTF), regionally known as
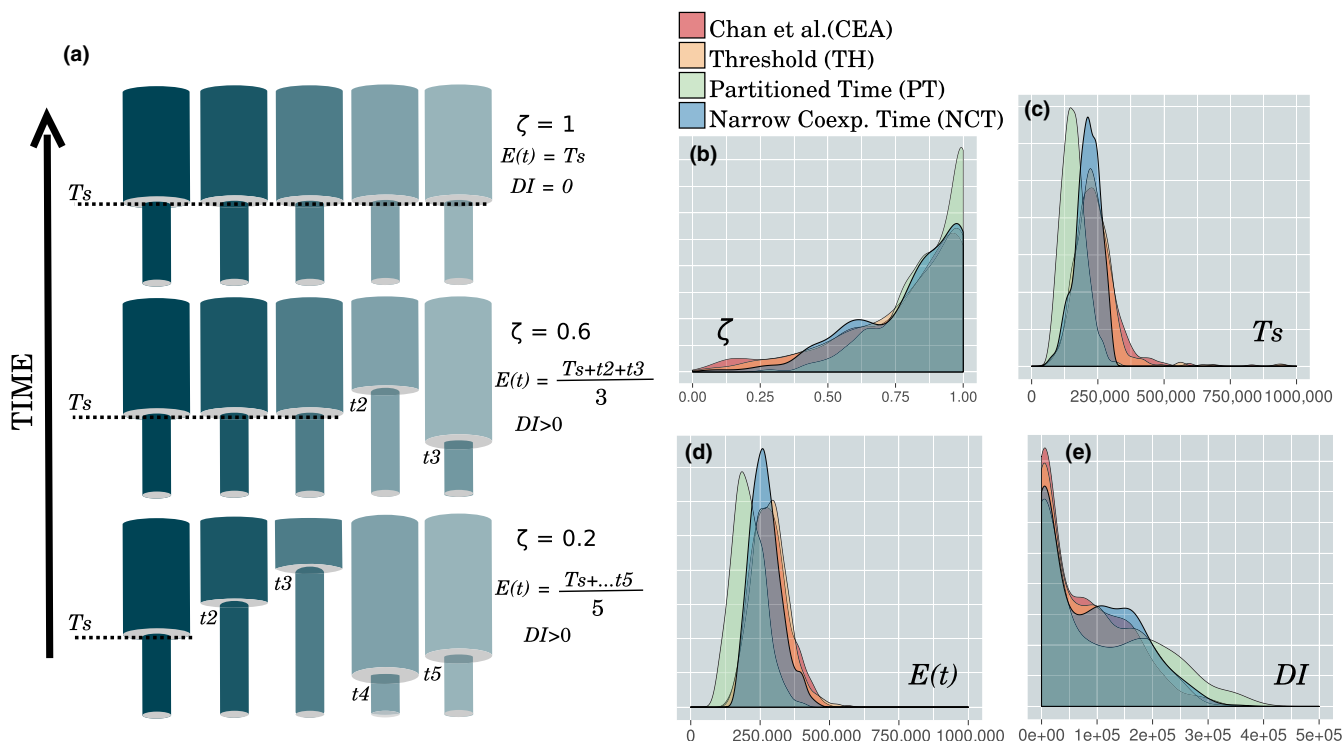
FIGURE 1 Schematic representation and estimates of hABC parameters: (a) three different ζ values for a five-population data set coexpansion model. (b–e) Superimposed posterior densities for each of the four parameters estimated under CET, TH, PT and NCT models using the *rejection* algorithm. (b) Proportion of coexpanding species (ζ), (c) coexpansion time (Ts), (d) average expansion time [$E(t)$] and (e) dispersion index [$DI = \text{var}(t)/E(t)$], where t are the expansion times [Colour figure can be viewed at wileyonlinelibrary.com]

Caatinga, characterized by xeric vegetation that tolerates strong seasonality and unpredictable droughts (Pennington, Lavin, & Oliveira-Filho, 2009; Pennington, Prado, & Pendry, 2000). Within the Caatinga, mesic forests occur as enclaves in areas subject to orographic rain (known as "*brejos-de-altitude*"). Eastwards of the Caatinga, along the coastal plain, wet forest is continuous, forming the Northern Atlantic Forest. The western edge of the Caatinga is isolated from the Amazon Forest by a stretch of savannah vegetation (Cerrado) (Figure 2). Paleoclimatic inferences suggest a history of climatic instability oscillating between moist and dry periods in the Caatinga over the last 1 million years (Auler et al., 2004a; Wang et al., 2007). Speleothem data indicate broad similarities in climatic history between Eastern Amazon and Northern Atlantic Forest in northeastern Brazil during Quaternary glacial–interglacial cycles, which may have favoured biotic exchanges between these two regions during climatic wet phases (Cheng et al., 2013). Phylogeographic analyses of wet forest-dwelling mammals, birds and reptiles support such exchanges during the Pleistocene (Batalha-Filho, Fjeldså, Fabre, & Miyaki, 2013; Costa, 2003; Prates et al., 2016), which implies the former presence of wet forest corridors in northeastern Brazil, where the Cerrado and Caatinga currently occur.

To test the hypothesis that the Caatinga has expanded over the last 1 million years, we analysed mitochondrial data of seven reptile and four amphibian species with different life histories from northeastern Brazil (Caatinga and transition zones). If the Caatinga

expanded during that time, most of its biota should show synchronous expansion, supporting the direct influence of past climate on demographic histories. In contrast, a pattern of idiosyncratic expansions, or the absence of demographic expansion for most species, suggests erratically expanding biomes and/or the predominance of species-specific ecological traits governing demographic changes, which is inconsistent with a major expansion of the Caatinga influencing the evolutionary history of the associated community.

2 | MATERIAL AND METHODS

2.1 | Hierarchical demographic models

Hierarchical demographic models can theoretically represent synchronous population size decreases (i.e., bottlenecks), increases (i.e., expansions) or a combination of both. Here, we evaluated the performance of the inference of simultaneous population expansions (i.e., coexpansion). To simulate coexpansion across populations, we implemented the hABC method developed by Chan et al. (2014), as well as three modified versions, within the R statistical environment (R Core Team 2016). The R code used in this study is included in the beta version of the PIPEMASTER R-package (Gehara, M.; Mazzocchini G.G. & Burbrink F.T. in prep.; www.github.com/gehara/PipeMaster). The method of Chan et al. (2014) allows the inference of four parameters: (i) the proportion of simultaneous demographic change (ζ); (ii) the timing of simultaneous change (Ts); (iii) the mean time of demographic changes
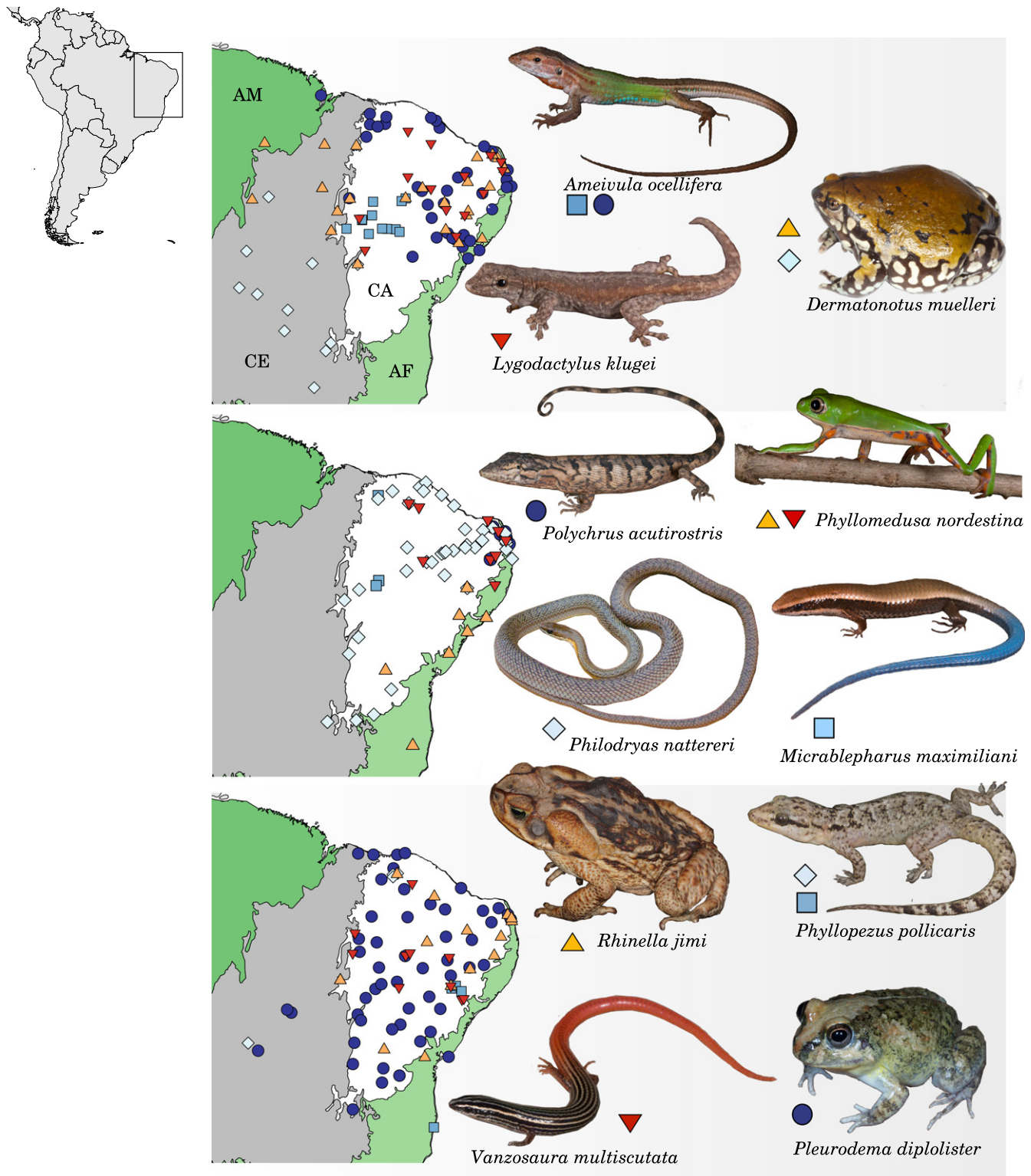
FIGURE 2 Maps showing northeastern Brazil with distribution maps of the 15 populations from seven reptile and four amphibian species analysed in this study. AM: Amazon, CE: Cerrado (savannah), CA: Caatinga (Seasonally Dry Tropical Forest), AF: Atlantic Forest [Colour figure can be viewed at wileyonlinelibrary.com]

[$E(t)$]; and (iv) the dispersion index (DI), or coefficient of variation of expansion times, that is the variance in expansion times divided by the mean [$E(t)$] (Figure 1a; see Chan et al., 2014). ζ and Ts are sampled from a uniform distribution, and when ζ is lower than one, population-specific expansion times (t_n) are sampled from the same prior (Figure 3a). Prior distributions of [$E(t)$] and DI are conditional to ζ and Ts. To simulate the magnitude of demographic change, current and ancestral population sizes (N_e) are sampled from population-specific priors.

Coalescent simulations are performed under those sampled parameters and hypersummary statistics (hSS) are calculated. Posterior distributions of the four parameters of the model are approximated by taking the observed hSS and applying an ABC rejection step on a collection of simulated hSS values.

The simulation phase consists of five steps:

1. Sample the hyperparameter ζ from a uniform prior distribution, with coexpansion proportions defined by the number of populations being simulated (e.g., for a 10-population data set, ζ will have 10 values from 0.1 to 1.0 with 0.1 increments); ζ defines only the number of coexpanding populations, the

specific populations that are expanding synchronously are randomly selected;

2. Draw a single expansion time for the coexpanding populations (Ts) and, if ζ is lower than 1.0, one expansion time for each non-coexpanding populations (t_n);

3. Draw population size parameters (current and ancestral) for each population from population-specific priors;

4. Sample mutation rates from population-specific priors;

5. Simulate genetic data under the coalescent model and calculate summary statistics.

In Chan et al.'s (2014) model (CEA hereafter), Ts and t are freely sampled from the same prior. Thus, simulations representing
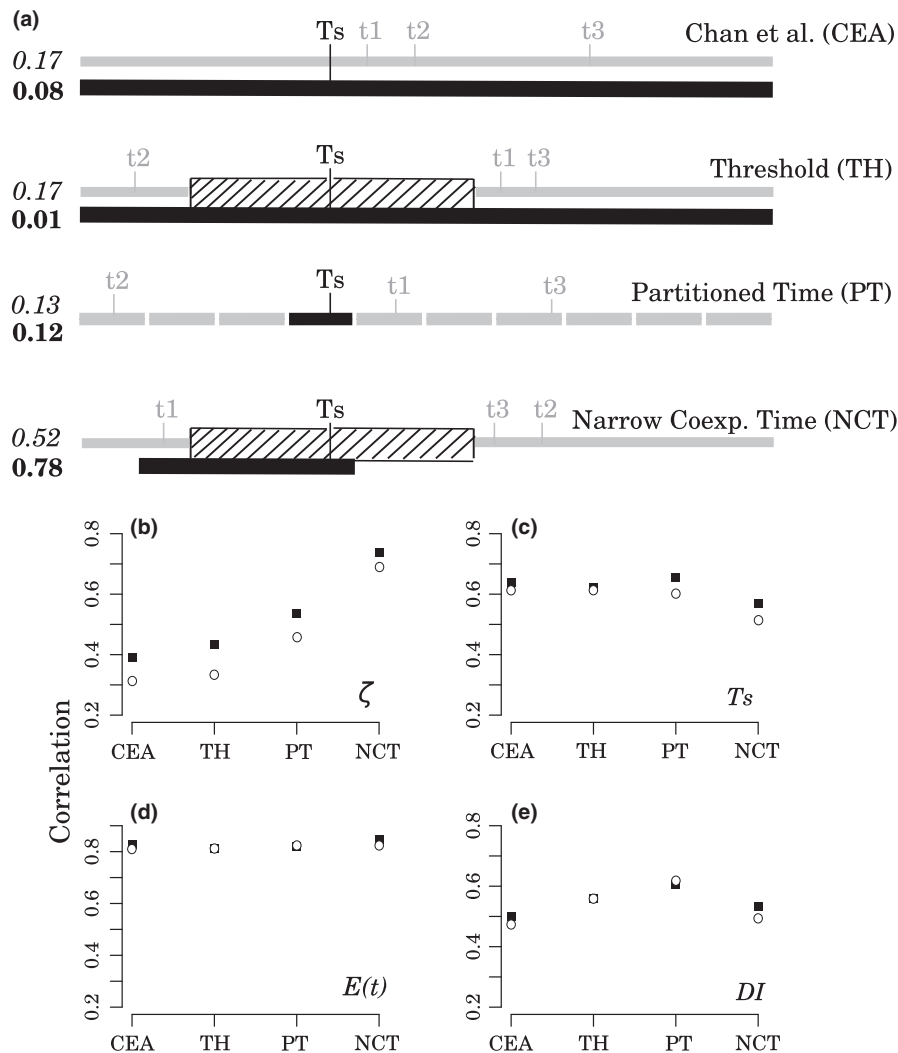

FIGURE 3 Schematic representation and comparison of the four hABC models: (a) expansion time priors and sampling procedure for the four coexpansion models simulated in this study. In this particular example, three species are not coexpanding in a 10 species data set. Black bars represent coexpansion time (Ts) prior distribution. Grey bars represent species-specific expansion time priors; and hashed boxes represent the threshold that forces Ts and t values to be n-years apart. Numbers on the left of each model show posterior probabilities of models for our empirical data using *neuralnet* (bold) and *rejection* (italic) algorithms. (b–e) correlations between 1,000 simulated pseudo-observed data (POD) values and estimated parameters using *neuralnet* and *rejection* algorithms under the four models. The ABC analyses was conducted using the *abc* package in the R statistical environment. Abbreviations on the x-axis represent the four models. Black squares indicate *rejection* algorithm and white circles indicate the *neuralnet* regression. (b) Proportion of coexpanding species (ζ); (c) coexpansion time (Ts); (d) average expansion time [$E(t)$]; (e) dispersion index [$DI = \mathrm{var}(t)/E(t)$, where t = expansion times]

idiosyncratic demographic changes will eventually have similar Ts and t_n values (Figure 3a). When Ts and t_n are similar, the derived hSS will be similar to those generated under higher ζ values. A large overlap between summary statistics generated under different ζ will potentially decrease the accuracy and precision of posterior estimates.

To increase confidence in estimates, we developed three modified versions of the CEA that incorporate different priors on the time parameters (Figure 3a). (i) The Threshold model (TH) differs from CEA by applying a threshold that forces the coexpansion time (Ts) and idiosyncratic expansion times (t_n) to be n-years apart, potentially increasing the difference among hSS from different ζ values (Figure 3a). This threshold is essentially the *buffer* parameter implemented in Multi-DICE when there is only a single coexpansion pulse (see Prates et al., 2016 and Xue & Hickerson, 2017). (ii) In the Partitioned Time model (PT), the expansion time (Ts) prior is equally subdivided into narrower, nonoverlapping blocks of priors that match the number of populations in the data set. In each simulation cycle, one of these blocks is randomly assigned to be the prior for the coexpansion time, while the remaining blocks are randomly assigned to any non-coexpanding species, so that each species expansion time and the coexpansion time are drawn from a different prior block (Figure 3a). This implementation ensures that expansion times are more evenly distributed across the expansion time prior, likely providing a better representation of asynchrony. (iii) The Narrow Coexpansion Time model (NCT), uses a threshold, as in TH, and a separate narrower prior for coexpansion time. When ζ is lower than one, the expansion time for the coexpanding species (Ts) is sampled from this narrower prior, and the expansion time for the non-coexpanding species (t), from a different wider prior (Figure 3a). This model represents a specific hypothesis of when the synchronous demographic change happened, narrowing possible Ts values, and potentially increasing precision and accuracy of ζ estimates by reducing the parameter space.

2.2 | Implementation and simulation of hierarchical models and hypersummary statistics

To simulate genetic data, we used *ms* (Hudson, 2002), which is implemented as a function inside the ʀ-package *phyclust* (Chen, 2011). We sampled population sizes from uniform distributions using the *runif* ʀ function (R Core Team 2016). Ancestral population sizes were defined as proportions of current sizes, while current effective population sizes were defined in number of individuals. Expansion times were defined in years.

We transformed sampled values to coalescent scale according to the *ms* manual as follows:

$$\Theta = 4N_e\mu \times \text{ bp } \times I$$

where μ is the sampled mutation rate per base pair per generation, bp is the number of base pairs, and I is the inheritance scalar (set as 0.25 for the mtDNA).

Expansion times were sampled in years and transformed to the coalescent scale as follows:

$$\tau = \frac{t/g}{4N_e \times I}$$

where t is the sampled expansion time in years, g is the generation time in years, and I is the inheritance scalar.

We selected a group of four summary statistics that are dependent on population size and demographic history (Tajima, 1989). These are similar to those used by Chan et al. (2014), but we replaced number of haplotypes by segregating sites (S). The number of segregating sites is generated automatically in *ms*, requiring no calculation and speeding up simulation cycles. The other summary statistics were nucleotide diversity (pi), haplotype diversity (h) and Tajima's D (TD). TD and pi were calculated using *ape* (Paradis, Claude, & Strimmer, 2004) and *pegas* (Paradis, 2010). Haplotype diversity was calculated using our own ʀ function according to Nei and Tajima (1981). We are aware of the influence of summary statistics on the power of ABC inferences (Burr & Skurikhin, 2013); however, an extensive evaluation of that influence was beyond the scope of our present study. The summary statistics used here should provide a good trade-off between demographic information and dimensionality (see Method evaluation below). To transform the summary statistics into hypersummary statistics, the four moments (mean, variance, kurtosis and skewness) for the distribution of each group of statistics were calculated using R-base functions or *e*1071 package (Meyer, Dimitriadou, Hornik, Weingessel, & Friedrich, 2015), totalling 16 hypersummary statistics.

2.3 | Method evaluation

To evaluate precision and accuracy of estimates of ζ, Ts, $E(t)$ and DI under the different coexpansion models, we used pseudo-observed data (POD) that consisted of simulated data for which the observed values of each parameter were known. We compared PODs with the respective parameter estimates using the ABC approach. We first performed, for each of the four models (CEA, TH, PT and NCT), 1,000,000 simulations of a data set containing 10 populations. We sampled the proportion of coexpanding populations from a uniform prior ranging from 0.1 to 1.0. For population-specific priors, we took a subset of the distributions used for the empirical data analysed in this study (Table 1). We used the same priors for ζ and Ts as used in our empirical data (see below). We then performed, for each model, an ABC parameter estimation for 1,000 PODs generated under the same model. We retained the 100 closest simulations using the *rejection* and Neural Network (*neuralnet*) algorithms implemented in *abc* (Csilléry, François, & Blum, 2012). Preliminary analyses with a larger number of simulations per model ($10\times$ more) did not result in higher precision and accuracy (data not shown). We plotted the POD parameters against estimated values for visual inspection and also performed Pearson's correlation tests (Pearson, 1895) to compare the performance of each model and each ABC algorithm. Additionally, to evaluate whether reduction of dimensionality of summary

TABLE 1 Priors distributions used in the simulations. Current N_e, ancestral N_e and expansion time were derived from IMa2 and Bayesian Skyline plot analyses. Mutation rates and generation times were taken from the literature (see Methods)

Species	N_e		Ancestral N_e (multiplier)		Expansion time		Mutation rate		bp	N	IS	Gen
	Min	Max	Min	Max	Min	Max	Mean	SD				
Ameivula ocellifera 1	1,500,000	4,000,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	370	58	0.25	2
Ameivula ocellifera 2	900,000	3,500,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	370	31	0.25	2
Dermatonotus muelleri 1[a]	120,000	1,200,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	514	16	0.25	1
Dermatonotus muelleri 2[a]	700,000	1,800,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	514	32	0.25	1
Lygodactylus klugei [a]	1,300,000	3,700,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	679	45	0.25	1
Philodryas nattereri	1,900,000	4,100,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	686	58	0.25	2
Phyllopezus pollicaris 1[a]	140,000	1,050,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	942	17	0.25	1
Phyllopezus pollicaris 2[a]	460,000	1,480,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	942	46	0.25	1
Microblepharus maximilliani [a]	870,000	2,400,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	1524	31	0.25	1
Vanzosaura multiscutata[a]	480,000	1,900,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	735	33	0.25	1
Phyllomedusa nordestina 1[a]	800,000	4,850,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	527	12	0.25	1
Phyllomedusa nordestina 2[a]	860,000	4,470,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	527	14	0.25	1
Rhinella jimi[a]	290,000	1,600,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	533	31	0.25	1
Pleurodema diplolister	4,000,000	7,020,000	0.001	0.1	50,000	1,000,000	1E-8	1.5E-9	601	166	0.25	1

bp, number of base pairs; N, number of samples; IS, inheritance scalar; Gen, Generation time.

[a]Priors used in the model evaluation.

statistics could increase the power of parameter estimates, we also performed a principal component analysis (PCA) on our summary statistics using the R-base prior to the ABC estimation. We compared the power of estimates using all 16 hypersummary statistics, the first three principal components (PCs), and the first 10 PCs of our summary statistics.

The four coexpansion models have different assumptions which might be violated by some data. To evaluate how each model performs under model violations, we compared estimates under a certain model (e.g., CEA) for PODs generated under the other three models (e.g., TH, PT and NCT). We conducted this exercise for all four models.

2.4 | Empirical data collection

We assembled mitochondrial gene fragments (mtDNA) from 11 species (Figure 2), three of them available on GenBank. We downloaded sequences of *Ameivula ocellifera*, *Phyllopezus pollicaris* and *Pleurodema diplolister*; obtained previously generated sequences of *Vanzosaura multiscutata* (Recoder et al., 2014); and generated new data for *Microblepharus maximiliani*, *Polychrus acutirostris*, *Lygodactylus klugei*, *Dermatonotus muelleri*, *Phyllomedusa nordestina*, *Rhinella jimi* and *Philodryas nattereri*. We isolated genomic DNA using standard salt extraction protocols (Bruford, Hanotte, Brookfield, Burke, & Hoelzel, 1992) or the Qiagen DNAeasy extraction kit and amplified mtDNA using PCR techniques (see Table S1 for primers and annealing temperatures). Sequence alignments with locality information are deposited in Dryad (https://doi.org/10.5061/dryad.789pv).

2.5 | Data treatment pipeline

Many single-locus data available online, particularly those originating from poorly studied areas of the globe, suffer from incomplete taxonomic and/or phylogeographic knowledge. These deficiencies may prevent comparative phylogeographic analyses or influence results because: (i) it is difficult to define a consistent prior distribution without phylogeographic knowledge, and prior distributions are crucial for any Bayesian analysis (Hickerson et al. 2013; Oaks et al. 2012); (ii) coalescent model used here assumes panmictic populations (Hudson, 2002) and the data may violate this assumption; (iii) the data may suggest demographic stability for a particular population. It is not logical, however, to test for synchronism of demographic change if the data does not support demographic change for the populations being analysed. To overcome these issues, we developed a data treatment pipeline to minimize potential pitfalls due to poorly illuminated data. We used the Bayesian Generalized Mixed Yule Coalescent (bGMYC; Fujisawa & Barraclough, 2013; Reid & Carstens, 2012) to define populations and ensure that the data contained the minimal populational substructure possible. We evaluated demographic change through time for each population using the Bayesian Skyline Plot method (BSP; Drummond, Rambaut, Shapiro, & Pybus, 2005) and neutrality tests. To generate prior information on long-term population sizes, we used IMa2 single-population models (Hey, 2010). These analyses were automated using R scripts that allowed easy data scalability. These scripts are available on GITHUB (github.com/gehara/ME_caatinga_scripts).

2.6 | Defining populations for Caatinga species with bGMYC

To exclude deep genetic structure from the data, we used the Bayesian Generalized Mixed Yule Coalescent method (bGMYC), which delimits species using single-locus data while accounting for topological uncertainty (Fujisawa & Barraclough, 2013; Reid & Carstens, 2012). The method sort the data based on the change from birth–death process (i.e., population-level process) to Yule process (i.e., above population-level process). Thus, we used the method to ensure that the data from each species contained only genetic information from population-level processes. For each species, we generated an ultrametric tree in BEAST v1.8.2 (Drummond, Suchard, Xie, & Rambaut, 2012) using the entire sequence alignment. Because the accuracy of gene tree topologies should have no major influence on our results, and they are not the main focus of the study, we assumed, for all species, an HKY+G substitution model with a Bayesian Skyline Plot (BSP) tree prior (Drummond et al., 2005). We used a strict molecular clock model with a substitution rate of 10^{-8}/site/generation for reptiles and amphibians (Crawford, 2003; Eo & DeWoody, 2010; Hauswaldt et al., 2014; Macey et al., 2001; Oliveira et al., 2015). We ran the MCMC for 20 million steps, sampling every 2,000 steps, and insured that the effective sample sizes of parameters were at least higher than 100 using CODA v0.19-1 (Plummer, Best, Cowles, & Vines, 2006). The resulting ultrametric trees were used as input for the BGMYC analyses. For these analyses, we ran an MCMC chain of 50,000 steps, with 40,000 steps of burn-in and thinned the results by 100. We ran the analyses using 100 trees randomly sampled from the population of 8,000 trees obtained from BEAST after a 20% burn-in. From the resulting BGMYC analyses, we retained all defined populations that contained at least 15 individuals. Only for *Polychrus acutirostris* and *Phyllomeduza nordestina* did we keep populations containing 11 individuals or more. We excluded populations with fewer individuals to avoid bias on demographic estimations.

2.7 | Population size changes through time and long-term N_e estimates

We evaluated population size changes through time using Bayesian Skyline Plot (BSP) analyses in BEAST v.1.8.2 (Drummond et al., 2005) and tested deviations from neutrality using Tajima's D (Tajima, 1989) and R_2 tests (Ramos-Onsins & Rozas, 2002). For the BSP, we used the same settings described above. To rescale the BSP, we multiplied the x-axis by the generation length of each species to obtain the time in years (Table 1). We calculated Tajima's D and R_2 and respective p-values using *pegas* with default settings. Populations that showed no signs of expansion according to the BSP and nonsignificant deviations from neutrality were excluded from the coexpansion simulations and were considered, a priori, as non-coexpanding. We also evaluated the long-term average population size using IMa2 (Hey, 2010). This analysis was run using a single-population model containing only the

theta parameter. We used a theta prior limit of 50 and assumed a mutation rate of 10^{-8}/site/generation, which was corrected to the gene mutation rate by multiplying by the number of base pairs of each respective gene. We took 10,000 samples with a burn-in of 200,000 genealogies and evaluated the effective sample sizes and chain mixing by checking the trend plots within the IMa2 output. The 95% high posterior densities of population sizes were used as boundaries for the current population size prior of each species.

2.8 | Priors' distributions and settings for the Caatinga hierarchical models

The species-specific priors for population sizes were derived from the BSP and the IMA2 analyses (Table 1). We set wide uniform priors from 50,000 to 1,000,000 years for the expansion times (Ts) for CEA, TH and PT models. According to the BSP results, we set for the NCT model an independent uniform prior from 50,000 to 300,000 years before present for Ts. For TH and NCT we applied a threshold of 50 ky of minimum difference between coexpansion time (Ts) and population-specific expansion times (t_n). Because different mitochondrial genes may have different mutation rates, we sampled mutation rates from a normal prior (mean: 10^{-8}/site/generation; SD: 1.5×10^{-9}) to accommodate this variation. A mean at 10^{-8}/site/generation is consistent across different studies that conducted mutation rate estimations for amphibians and reptiles (Crawford, 2003; Eo & DeWoody, 2010; Hauswaldt et al., 2014; Macey et al., 2001; Oliveira et al., 2015). We assumed a generation length of 1 year for amphibians and 2 years for reptiles (Dessauer, Cole, & Townsend, 2000; Duellman & Trueb, 1994; pp. 33). For each model, we performed 11,200,000 simulations, corresponding to 800,000 data sets per species. All simulations were carried out in the City University of New York High Performance Computing Center; however, it is important to note that to simulate the same amount of data in a single core with 2.6 GHz of processing power would take only ~10 days. We evaluated the fit of the simulations to our observed data by performing a principal component analysis (PCA) on the hypersummary statistics, plotting the first and second PCs, and the first and third PCs, of simulated and observed data (Fig. S1).

2.9 | Model selection and parameter estimation

We conducted an ABC model selection to evaluate which of the four simulated models better fit our data. For that we used 2×10^6 simulations of each model. We applied both the *rejection* and the *neuralnet* algorithms with *sizenet* of 10, *numnet* of 20 and maximum iterations of 2,000, without correction for heteroscedasticity, and with a tolerance of 0.00005. We also estimated, for each one of the four simulated models, the posterior distribution of each parameter for our empirical data using *abc* function (Csilléry et al., 2012) with the same settings as above and a tolerance of ~0.00004.

3 | RESULTS

3.1 | Method evaluation

All estimated parameters (ζ, $E(t)$, Ts and DI) under all models were significantly correlated with true POD values ($p < 10^{-5}$), using both *rejection* and *neuralnet* regression algorithms (Table 2; Figure 3b–e). Nevertheless, there was a large variation on precision and accuracy across models for estimates of the proportion of coexpanding populations (ζ) (between 0.31 and 0.74). For both accuracy and precision among all models, NCT performed the best for estimating ζ (Figure 4a,b), while CET had the poorest performance for this parameter (Figure 4a,b). In general, under CET, TH and PT, high ζ values were underestimated while low ζ values were overestimated, generating a distortion on posteriors, particularly when using the *neuralnet* algorithm. Distortion on posterior distributions may happen when the true value is at the boundary of a uniform prior (Beaumont, 2011).

Among the other parameters, the mean expansion time [$E(t)$] showed the highest correlation across all models ($r > .8$; Figures 3d and 5a,b). $E(t)$ is the easiest parameter to estimate and can be obtained with the same confidence under all models (Figure 5a,b). The dispersion index (DI) showed the lowest correlation ($.45 < r < .6$), and the coexpansion time showed intermediate values ($.5 < r < .67$; Figure 3c,e). Parameterizations under PT showed slightly higher correlations for time of simultaneous demographic change estimates (Ts; Figure 4c,d) and for the dispersion index (DI; Figure 5c,d).

The reduction of summary statistics dimensions using PCA did not improve posterior estimates. Estimates using the first three PCs were less accurate and precise under all models, as indicated by lower correlation values. Estimates using the first 10 PCs or all 16 hypersummary statistics were similarly correlated, suggesting comparable accuracy and precision of estimates (Fig. S2).

The NCT model was the most sensitive to violations and showed poor performance when PODs were generated under a different model (Fig. S6). All models yielded better estimates of ζ when PODs were generated under NCT, and worse estimates of ζ when PODs were generated under CEA (Figs S3–S6). The average expansion time, $E(t)$, was the least sensitive parameter, showing higher correlation even under model violations.

3.2 | Population sizes change and long-term N_e estimates for the Caatinga community

After delineating lineages showing deep genetic breaks with bGMYC, we defined 15 populations useful for further testing. The following species were divided into two populations: *Ameivula ocellifera*, *Dermatonotus muelleri*, *Phyllopezus pollicaris* and *Phyllomedusa nordestina*.

TABLE 2 Pearson's correlation between PODs and parameters estimated with *rejection* and *neuralnet* ABC algorithms

	Rejection			Neuralnet	
	rho	CI		rho	CI
zeta					
CEA	0.389848	0.3359745–0.4411785		0.312631	0.2555915–0.3675016
TH	0.434682	0.3830099–0.4836422		0.333775	0.2775244–0.3877450
PT	0.537105	0.4914760–0.5797924		0.4574406	0.4069890–0.5051097
NCT	0.737049	0.7073769–0.7641273		0.6905998	0.6567226–0.7216954
Ts					
CEA	0.637605	0.5993004–0.6729964		0.6130591	0.5728370–0.6503359
TH	0.622318	0.5828096–0.6588916		0.6111111	0.5707403–0.6485346
PT	0.65361	0.6166016–0.6877368		0.6021918	0.5611473–0.6402821
NCT	0.570439	0.5270855–0.6108312		0.5149131	0.4678544–0.5590604
E(t)					
CEA	0.826121	0.8053745–0.8448466		0.8094157	0.7869082–0.8297723
TH	0.813968	0.7919361–0.8338828		0.8146041	0.7926391–0.8344572
PT	0.819376	0.7979137–0.8387637		0.822642	0.8015252–0.8417095
NCT	0.845919	0.8273115–0.8626729		0.8261054	0.8053570–0.8448323
DI					
CEA	0.499194	0.4511632–0.5443419		0.4724239	0.4228137–0.5192108
TH	0.560024	0.5159433–0.6011468		0.5595516	0.5154382–0.6007071
PT	0.604872	0.5640291–0.6427631		0.6190278	0.5792642–0.6558523
NCT	0.53402	0.4881885–0.5769138		0.4940027	0.4456577–0.5394745

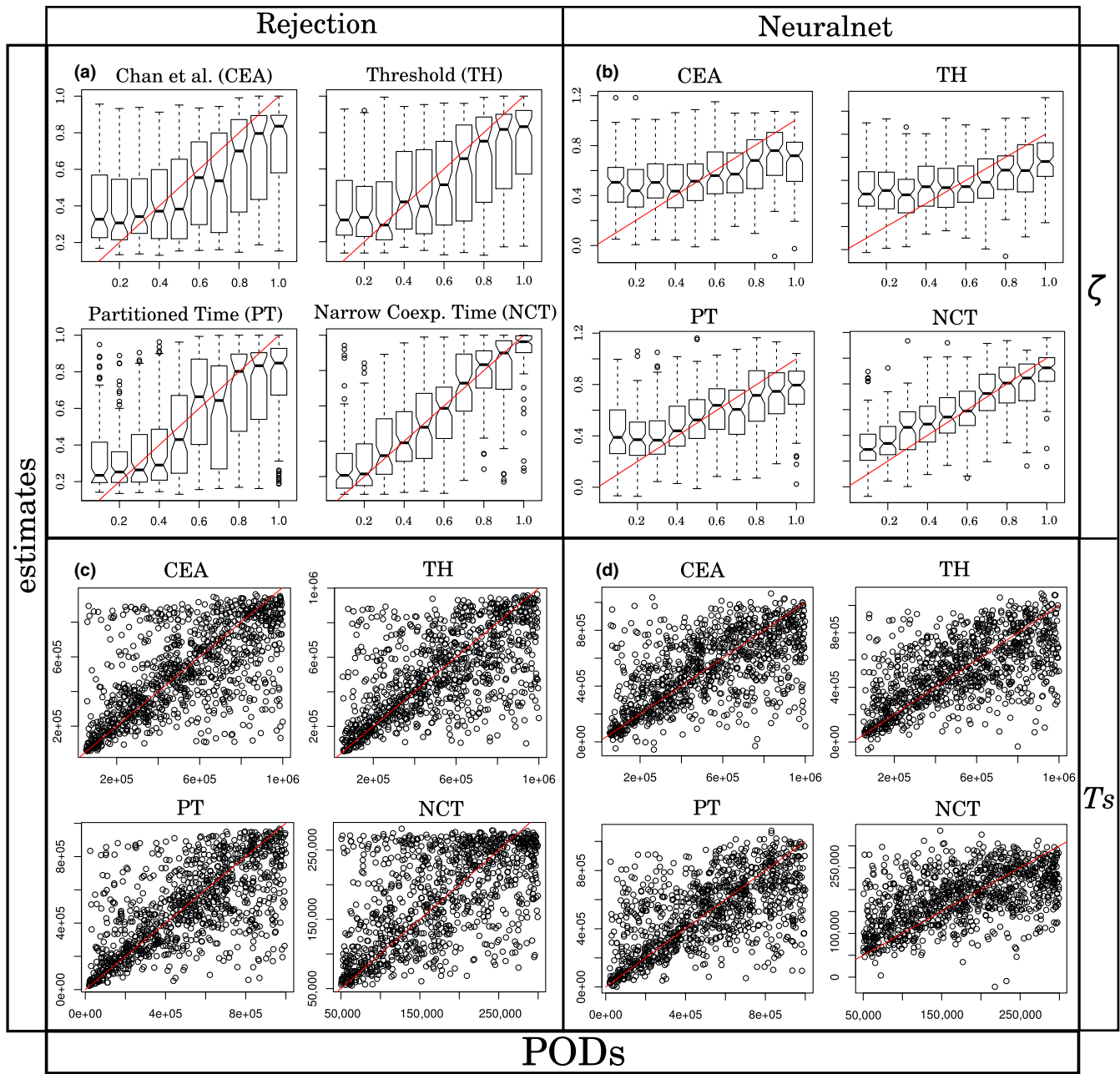All correlations are significant, $p < .001$.

FIGURE 4 Plots of 1,000 PODs and their respective estimates showing the precision and accuracy of: (a, b) proportion of synchronous population size changes (ζ); (c, d) time of synchronous expansion (Ts). (a, c) Estimates with *rejection*. (b, d) Estimates with the *neuralnet*. Red line represents a perfect correlation between estimate and true value [Colour figure can be viewed at wileyonlinelibrary.com]

Of the 15 populations, 14 showed signal of expansion as suggested by BSPs (Figure 6). All populations showed negative Tajima's D but only three had significant values, while R_2 values were significant for 11 populations (Table 3). *Polychrus acutirostris* showed population stability in the BSP and nonsignificant R_2 and Tajima's D and was therefore excluded from the coexpansion simulations. BSPs suggested that most populations expanded between 50 and 300 ky. The two populations of *A. ocellifera*, *Philodryas nattereri* and *P. nordestina* showed earlier expansion (Figure 6). Nevertheless, all patterns of expansion fell within the Pleistocene. Long-term estimates of N_e for each population ranged from ~418,000

(*Dermatonotus muelleri* population 1) to ~5,000,000 (*Pleurodema diplolister*; see Table 1 for all confidence intervals), as suggested by IMa2 single-population analyses.

3.3 | Parameter estimates and model selection for the Caatinga community

Estimates were largely congruent among models and generally indicated highly synchronous population expansion occurring in the late Pleistocene. Under all models, posterior distributions obtained with the *rejection* algorithm showed high density around high values of ζ,
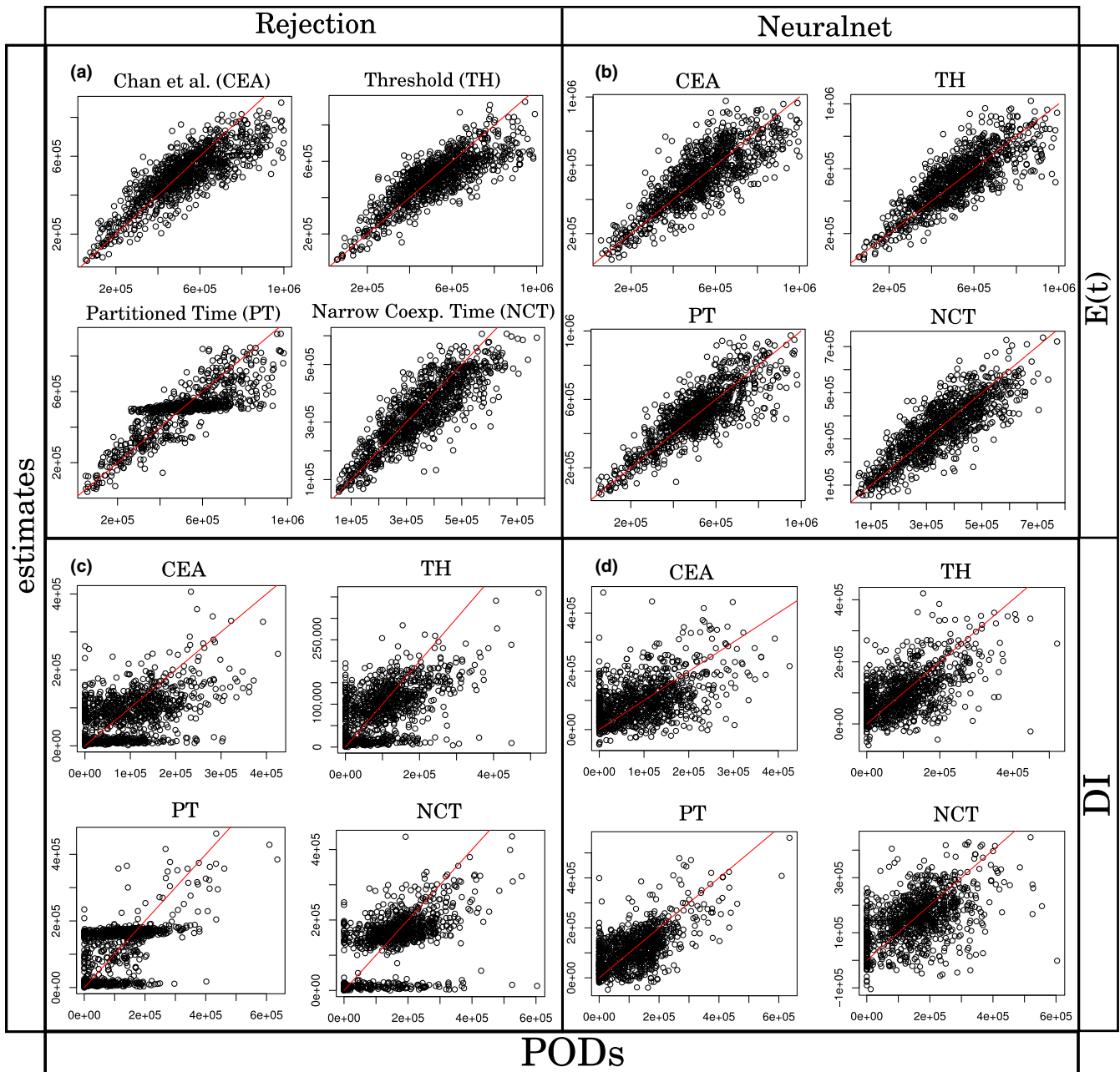
FIGURE 5 Plots of 1,000 PODs and their respective estimates showing the precision and accuracy of: (a, b) mean expansion time [$E(t)$]; (c, d) dispersion index (DI). (a, c) Estimates with *rejection*. (b, d) Estimates with the *neuralnet*. Red line represents a perfect correlation between estimate and true value [Colour figure can be viewed at wileyonlinelibrary.com]

largely overlapping Ts and $E(t)$ distributions, and DI posteriors with high density near zero (Figure 1b–e, Table 4). The *neuralnet* regression produced relatively lower ζ under all models (Fig. S7). The mode of the posterior density for the proportion of coexpanding populations (ζ) ranged from 0.68 (under CEA and *neuralnet* algorithm) to 0.99 (under PT and *rejection* algorithm; Table 4, Figure 1b). The posterior distribution of the coexpansion time (Ts) differed among models with modes ranging between 118 and 224 ky (Table 4, Figure 1c). The modes of average expansion time [$E(t)$] ranged from 166 to 294 ky, depending on the model and on the ABC algorithm used (Table 4, Figure 1d). The modes of dispersion index (DI) ranged

from 4 to 127 ky (Figure 1e). Model selection indicated NCT as having the highest probability, with 0.52 probability using *rejection* and 0.78 using the *neuralnet* (Figure 3a).

4 | DISCUSSION

4.1 | Model evaluation

The most accurate way to estimate the level of demographic synchronism (ζ) is by providing an a priori hypothesis for the timing of synchronous change (Ts), as suggested by the better performance of
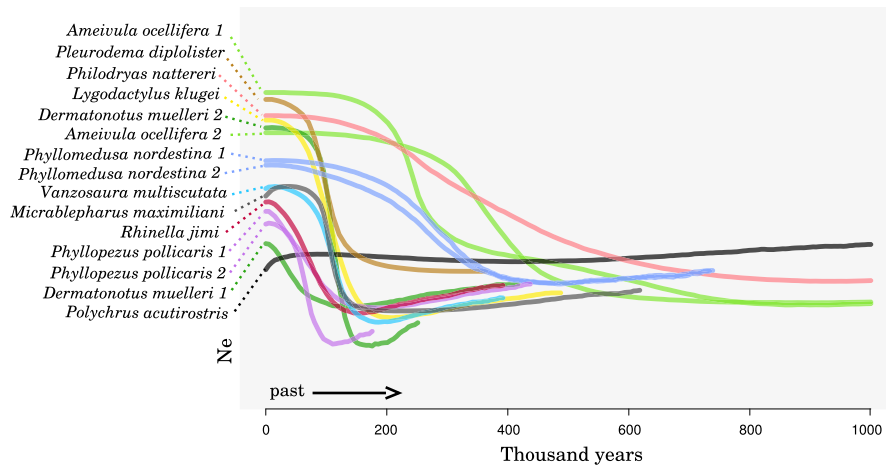
FIGURE 6 Combined Skyline plots median estimates of population size change through time for all 15 populations. The *x*-axis represents time, with the present at the origin. The *y*-axis represents effective population size [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Neutrality tests with significance values for all population analysed in this study

	R_2	*p*-value	Tajima's D	*p*-value
Ameivula ocellifera 1	0.04	.021	−1.71	.087
Ameivula ocellifera 2	0.07	.079	−1.07	.286
Dermatonotus muelleri 1	0.12	.264	−0.52	.603
Dermatonotus muelleri 2	0.04	.000	−2.41	.016
Lygodactylus klugei	0.03	.000	−2.23	.025
Microblepharus maximiliani	0.05	.001	−1.94	.052
Polychrus acutirostris	0.18	.741	−1.14	.252
Pleurodema diplolister	0.02	.008	−2.17	.030
Philodryas nattereri	0.04	.010	−1.82	.069
Phyllomedusa nordestina 1	0.08	.002	−1.52	.128
Phyllomedusa nordestina 2	0.07	.000	−1.36	.174
Phyllopezus pollicaris 1	0.08	.017	−1.89	.058
Phyllopezus pollicaris 2	0.06	.064	−1.44	.149
Rhinella jimi	0.06	.017	−1.71	.087
Vanzosaura multiscutata	0.05	.003	−1.77	.076

NCT model. Our results also suggest that it is hard to estimate fine temporal differences in expansion times. All models performed better for PODs generated under the NCT, indicating that it is easier to

estimate ζ when the empirical data show large temporal differences between the time of the simultaneous demographic change (Ts) and idiosyncratic expansion times (t_n).

Defining an a priori hypothesis for the timing of synchronous change, with a narrow prior on *Ts*, may generate inconsistency between empirical and simulated data. This inconsistency may explain the high sensitivity of NCT to model violations. Thus, it is essential to ensure that prior distributions of hABC models are consistent with the data. Furthermore, defining informative priors in hABC is crucial because wide, less informative uniform priors may bias results towards high levels of synchronism (Hickerson et al. 2013; Oaks et al. 2012). Our use of narrow priors and our alternative parameterizations may explain the lack of such bias in our model evaluation (Hickerson et al. 2013).

Forcing expansion times to be more evenly distributed across the time prior (PT model) can marginally improve estimates of the coexpansion time (*Ts*), the mean expansion time [*E*(*t*)] and the dispersion index (*DI*). One advantage of PT over NCT is that no a priori hypothesis for the time of synchronous demographic change (*Ts*) is assumed. Thus, estimates of *Ts* under PT should help define *Ts* priors for the NCT model, which is more accurate for estimating ζ.

Our method evaluation was tested for our case study, so generalizations should be made with care. Still, our conclusions should

TABLE 4 Mode and credible intervals of posterior distributions of parameters estimated under each of the four models using the *neuralnet* (italic) and the *rejection* algorithm (bold)

	zeta		Ts		E(t)		DI	
	neuralnet	**rejection**	*neuralnet*	**rejection**	*neuralnet*	**rejection**	*neuralnet*	**rejection**
CEA	*0.79 (0.36–0.94)*	**0.96 (0.14–1)**	*212 (104–324)*	**224 (117–434)**	*274 (200–362)*	**270 (174–432)**	*112 (79–185)*	**6 (0–236)**
TH	*0.68 (0.4–0.8)*	**0.97 (0.21–1)**	*204 (113–327)*	**221 (123–406)**	*297 (233–386)*	**295 (172–428)**	*125 (86–229)*	**5 (0–244)**
PT	*0.95 (0.6–1.05)*	**0.99 (0.5–1)**	*118 (73–184)*	**147 (85–263)**	*166 (102–240)*	**185 (110–323)**	*59 (10–150)*	**4 (0–332)**
NCT	*0.82 (0.45–0.93)*	**0.98 (0.4–1)**	*159 (111–238)*	**218 (110–287)**	*264 (200–335)*	**262 (176–386)**	*115 (81–202)*	**4 (0–244)**

Posterior distributions were obtained with the ABC *rejection* and *neuralnet* regression algorithm using the *abc* R-package and retaining the closest 500 simulations. Between parentheses is the 2.5–97.5% HDP. Values of *Ts*, *E*(*t*) and *DI* are shown in Kyears.

provide a guide for future analysis. Future studies should consider at least PT and NCT models to confirm posterior estimates. A model selection step can also indicate the best model before calculating posterior distributions. Also, as in any ABC analysis, it is important to assess model fit using, for instance, principal component analysis and/or cross-validation experiments (Cornuet, Ravigne, & Estoup, 2010).

Even with good model fit, posterior estimates should be interpreted cautiously and should provide an idea of how strong the genetic signal is in support of temporal concordance or discordance among independent demographic histories. Single-locus data have relatively more noise due to the stochasticity of the coalescent process, which is likely the reason for the wide credible intervals of posteriors. It is difficult to estimate all four parameters of interest here with high precision and accuracy. Combining results from posteriors of all parameters should help to properly evaluate the signal of demographic synchronism. For instance, when ζ equals one, $E(t)$ and Ts have the same value, and DI equals zero (Figure 1a). Thus, high similarity between $E(t)$ and Ts posterior distributions and high density of DI posteriors near zero are all indicative of highly synchronous demographic changes.

A potential improvement to the method could include models that simulate pulses of demographic changes similar to pulses of divergence simulated in msBayes or the demographic pulses of Multi-DICE (Hickerson et al., 2007; Xue & Hickerson, 2017). These models could represent alternative scenarios that may fit the data better, particularly when synchronism is not high, and could be used to gain refined insights on the demographic history of a community. However, the level of confidence on the inference of demographic pulses using single-locus data is unknown and should be evaluated by future studies.

4.2 | Simultaneous expansion of the Caatinga herpetofauna

We show that the Caatinga herpetofauna is characterized by synchronous population expansions around the beginning of the Late Pleistocene (Modes across models: 118–224 ky). There is, under all models, a high proportion of coexpanding populations (high ζ), largely overlapping posterior distribution of coexpansion times (Ts) and average expansion times [$E(t)$], and high density of dispersion indexes (DI) close to zero. According to BSPs and neutrality tests, only one species, *Polychrus acutirostris*, shows population stability and by definition does not have a synchronous history with the other species.

Our results support a major spatial reorganization of Caatinga landscape in the late Pleistocene, with expansion of dry formations and related contraction of wet forests in northeastern Brazil. Several palaeontological studies show dynamic climatic changes in northeastern Brazil for the last 1 Mya, with periods of aridity and changes in the Caatinga vegetation (Auler & Smart, 2001; Auler et al., 2004b; De Oliveira, Magno, & Suguio, 1999; Wang et al., 2004). Many phylogeographic studies show evidence of population and/or range expansion for Caatinga species (Caetano et al., 2008; Faria,

Nascimento, De Oliveira, & Bonvicino, 2013; Franco & Manfrin, 2013; Magalhaes et al., 2014; Monteiro, Donnelly, Beard, & Costa, 2004; Oliveira et al., 2015; Thomé et al., 2016; Werneck, Leite, Geurgas, & Rodrigues, 2015), although our study uniquely models historical demography at the community level using hABC.

We analysed species that are phylogenetically and ecologically distinct, including some taxa with high seasonal activity, such as *Dermatonotus muelleri* (Nomura, Rossa-Feres, & Langeani, 2009), and others that are active year round, like *Ameivula ocellifera* (Mesquita & Colli, 2003). Despite these ecological differences, most taxa showed population expansion. In such cases, a hypothesis of climatically driven demographic change requires fewer assumptions by assigning the cause of expansion to a single factor. It is therefore unexpected to have synchronous expansion of phylogenetically distant species driven by species-specific ecological factors. Still, our target species are characterized by very different ecological adaptations and life history strategies that maintain large population sizes in the face of climatic fluctuations, increasing aridity and seasonality. Thus, the expansion of Caatinga habitats was likely extensive, affecting synchronously increasing the population sizes of distantly related dry-adapted organisms.

Fossil records from late Pleistocene caves from northeastern Brazil show the presence of arboreal and wet forest-dwelling mammal species that are no longer present in the area, such as extinct giant monkeys and bat species currently restricted to wet forests (Auler et al., 2004b; Cartelle & Hartwig, 1996; Czaplewski & Cartelle, 1998). This suggests that increased aridity likely decreased habitat suitability for wet forest-adapted species, leading to population contraction and local extinction. On the other hand, species with traits allowing them to expand their population sizes could remain in the area, forming communities adapted to aridity and seasonality. For example, two species of frogs analysed here, *Dermatonotus muelleri* and *Pleurodema diplolister*, exhibit aestivation behaviour, which is a clear adaptation to aridity (Navas & Carvalho, 2010).

We found synchronous population expansion of codistributed Caatinga species at ~150 ky (2.5%ile: 85 ky; 97.5%ile: 260 ky), based on estimates under PT and *rejection* algorithm, and at ~220 ky (2.5%ile: 110 ky; 97.5%ile: 287 ky) under NCT and *rejection* algorithm. During this time, three short wet phases (~200 to ~210 ky; ~175 to ~185 ky; ~130 to ~140 ky) and three longer dry phases (~185 to ~200 ky; ~140 to ~175 ky; ~110 to ~130 ky) occurred in the Caatinga (Auler et al., 2004a; Wang et al., 2004). It is difficult to associate the credible intervals of Ts estimates with a specific inferred climatic phase, but they still indicate that the size of the Caatinga increased in the last ~260 ky. Also in support of that is the highest probability found for the NCT model, which assumes a Ts between 50 and 300 Ky. In contrast to the expansion of Caatinga and dry-adapted taxa, population contraction is expected for species restricted to wet forest enclaves and adjacent continuous wet forest from northeastern Brazil. Interestingly, recent studies suggest population contraction of wet forest-dwelling reptiles and birds in Northern Atlantic Forest (NAF) and Eastern Amazonia (EAM; Cabanne et al., 2016; Prates et al., 2016).

Paleoclimatic inferences suggest that EAM and NAF were less stable than the southern Atlantic Forest (SAF) and western Amazon (WAM). They also suggest that when the NAF and EAM were more arid, the SAF and WAM were more humid (Cheng et al., 2013). The expansion of dry biomes with fragmentation of wet forest in northeastern Brazil and the evidence of southward displacement of frogs and birds from the Atlantic Forest are in agreement with this hypothesis (Cabanne et al., 2016; Carnaval et al., 2009; Menezes et al., 2016). Considering more recent times, from the Last Glacial Maximum to current conditions, different paleoclimatic distribution models support either stability (Werneck, Costa, Colli, Prado, & Sites, 2011) or instability (Collevatti et al., 2013) in parts of the Caatinga. However, the occurrence of population expansion for the majority of species undermines the hypothesis of a "Caatinga nucleus" representing a stable refuge of STDF throughout the Pleistocene (see also de Melo, Lima-Ribeiro, Terribile, & Collevatti, 2016).

Our results support the hypothesis that wet forest enclaves found in moister highlands ("*brejos-de-altitude*") represent micro-refuges of a once more widespread wet forest that likely connected the Amazon and the Atlantic Forest during the Pleistocene (Andrade-Lima, 1982; Batalha-Filho et al., 2013; Costa, 2003; Melo Santos, Cavalcanti, da Silva, & Tabarelli, 2007; Prado & Gibbs, 1993; Prates et al., 2016). Herpetological inventories also support this "forest connectivity" hypothesis and show higher species richness in humid highland areas (Guedes, Nogueira, & Marques, 2014; Loebmann & Haddad, 2010; Pereira Filho & Montingelli, 2011; Ribeiro, Roberto, Sales, Ávila, & de Oliveira Almeida, 2012).

We present evidence in support of Late Pleistocene expansion of the Caatinga and adjacent dry formations in northeastern Brazil. This overwhelming signal strongly indicates that populations have expanded at congruent times regardless of their heterogeneous physiological and ecological requirements. However, to confirm our findings, future studies should investigate synchronous population expansion of other groups of Caatinga organisms, and also population contraction of wet forest-adapted species in the same area. With the use of genomic data, it should be easier to estimate narrower credible intervals and factor out selection from demographic change to better determine whether simultaneous expansion is associated with convergent adaptation in the formation of Caatinga biota.

DATA ACCESSIBILITY

Sequence alignments with locality information are deposited in Dryad (https://doi.org/10.5061/dryad.789pv). Code used for the coexpansion simulations are available in www.github.com/gehara/PipeMaster. Pipeline for data treatment is available in GITHUB (www.github.com/gehara/ME_caatinga_scripts).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M.G., A.A.G. and F.T.B. designed the research; A.A.G., F.P.W., E.F.O., E.M.F., F.C., F.M.M., F.M.L., R.M., R.S.F. and V.A.S.P. generated the sequence data; M.G., A.A.G., F.P.W., J.W.S. Jr., G.R.C., G.C.C. and F.T.B. contributed reagents/materials/analysis tools; M.G. and F.T.B. analysed data; M.G. and F.T.B. wrote the study and all authors helped improving the final version.

REFERENCES

Andrade-Lima, D. (1982). Present-day forest refuges in Northeastern Brazil. In G. Prance (Ed.), *Biological diversification in the tropics* (pp. 245–251). New York: Columbia University Press.

Auler, A. S., & Smart, P. L. (2001). Late Quaternary paleoclimate in semi-arid northeastern Brazil from u-series dating of travertine and water-table speleothems. *Quaternary Research*, 55, 159–167.

Auler, A. S., Wang, X., Edwards, R. L., Cheng, H., Cristalli, P. S., Smart, P. L., & Richards, D. A. (2004a). Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. *Journal of Quaternary Science*, 19, 693–701.

Auler, A. S., Wang, X., Edwards, R. L., Cheng, H., Cristalli, P. S., Smart, P. L., & Richards, D. A. (2004b). Palaeoenvironments in semi-arid northeastern Brazil inferred from high precision mass spectrometric speleothem and travertine ages and the dynamics of South American rainforests. *Speleogenesis and Evolution of Karst Aquifers*, 2, 1–4.

Batalha-Filho, H., Fjeldså, J., Fabre, P. H., & Miyaki, C. Y. (2013). Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *Journal of Ornithology*, 154, 41–50.

Beaumont, M. A. (2011). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41, 379–406.

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, 41, 8–22.

Bruford, M. W., Hanotte, O., Brookfield, J. F. Y., Burke, T., & Hoelzel, A. R. (1992). Single-locus and multilocus DNA fingerprint. In A. R. Hoelzel (Ed.), In: *Molecular genetic analysis of populations: A practical approach* (pp. 225–270). Oxford: IRL Press.

Burbrink, F. T., Chan, Y. L., Myers, E. A., et al. (2016). Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. *Ecology Letters*, 19, 1457–1467.

Burr, T., & Skurikhin, A. (2013). Selecting summary statistics in approximate Bayesian computation for calibrating stochastic models. *BioMed Research International*, 2013, 1–10.

Cabanne, G. S., Calderón, L., Trujillo-Arias, N., et al. (2016). Pleistocene climate changes affected species ranges and evolutionary processes in the Atlantic Forest. *Biological Journal of the Linnean Society*, 199, 856–872.

Caetano, S., Prado, D., Pennington, R. T., et al. (2008). The history of Seasonally Dry Tropical Forests in eastern South America: Inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, 17, 3147–3159.

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science (New York, N.Y.)*, 323, 785–789.

Cartelle, C., & Hartwig, C. (1996). A new extinct primate among the Pleistocene megafauna of Bahia, Brazil. *Proceedings of the National Academy of Sciences*, 93, 6405–6409.

Chan, Y. L., Schanzenbach, D., & Hickerson, M. J. (2014). Detecting concerted demographic response across community assemblages using hierarchical approximate Bayesian computation. *Molecular Biology and Evolution*, 31, 2501–2515.

Chen, W. C. (2011). *Overlapping codon model, phylogenetic clustering, and alternative partial expectation conditional maximization algorithm*. Ames, IA: Iowa State University.

Cheng, H., Sinha, A., Cruz, F. W., et al. (2013). Climate change patterns in Amazonia and biodiversity. *Nature Communications*, 4, 1411.

Collevatti, R. G., Terribile, L. C., de Oliveira, G., et al. (2013). Drawbacks to palaeodistribution modelling: The case of South American seasonally dry forests. *Journal of Biogeography*, 40, 345–358.

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

Costa, L. P. (2003). The historical bridge between the Amazon and the Atlantic Forest of Brazil: A study of molecular phylogeography with small mammals. *Journal of Biogeography*, 30, 71–86.

Crawford, A. J. (2003). Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology*, 12, 2525–2540.

Csilléry, K., François, O., & Blum, M. G. B. (2012). Abc: An R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, 3, 475–479.

Czaplewski, N. J., & Cartelle, C. (1998). Pleistocene bats from cave deposits in Bahia, Brazil. *Journal of Mammalogy*, 79, 784–803.

De Oliveira, P. E., Magno, A., & Suguio, K. (1999). Late Pleistocene/Holocene climatic and vegetational history of the Brazilian Caatinga: The fossil dunes of the middle São Francisco River. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 152, 319–337.

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

Dessauer, H. C., Cole, C. J., & Townsend, C. R. (2000). Hybridization among western whiptail lizards (*Cnemidophorus tigris*) in Southwestern New Mexico: Population genetics, morphology, and ecology In three contact zones. *Bulletin of the American Museum of Natural History*, 246, 1–148.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185–1192.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29, 1969–1973.

Duellman, W. E., & Trueb, L. (1994). *Biology of amphibians* (pp. 670). New York: McGrawHill Inc.

Eo, S. H., & DeWoody, J. A. (2010). Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. *Proceedings of the Royal Society of London B: Biological Sciences*, 277, 3587–3592.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905.

Faria, M. B., Nascimento, F. F., De Oliveira, J. A., & Bonvicino, C. R. (2013). Biogeographic determinants of genetic diversification in the mouse opossum *Gracilinanus agilis* (Didelphimorphia: Didelphidae). *Journal of Heredity*, 104, 613–626.

Franco, F. F., & Manfrin, M. H. (2013). Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. *Journal of Biogeography*, 40, 142–154.

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: A revised method and evaluation on simulated datasets. *Systematic Biology*, 62, 707–724.

Garrick, R. C., Bonatelli, I. A. S., Hyseni, C., et al. (2015). The evolution of phylogeographic datasets. *Molecular Ecology*, 24, 1164–1171.

Graham, C. H., Moritz, C., & Williams, S. E. (2006). Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy Sciences*, 103, 632–636.

Guedes, T. B., Nogueira, C., & Marques, O. A. V. (2014). Diversity, natural history, and geographic distribution of snakes in the Caatinga, Northeastern Brazil. *Zootaxa*, 3863, 1–93.

Hauswaldt, J. S., Angelini, C., Gehara, M., et al. (2014). From species divergence to population structure: A multimarker approach on the most basal lineage of Salamandridae, the spectacled salamanders (genus *Salamandrina*) from Italy. *Molecular Phylogenetics and Evolution*, 70, 1–12.

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

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

Hickerson, M. J., Stahl, E. A., & Lessios, H. A. (2006). Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, 60, 2435–2453.

Hickerson, M. J., Stahl, E., & Takebayashi, N. (2007). msBayes: Pipeline for testing comparative phylogeographic histories using hierarchical approximate Bayesian computation. *BMC Bioinformatics*, 8, 268.

Hickerson, M. J., Stone, G. N., Lohse, K., Demos, T. C., Xie, X., Landerer, C., & Takebayashi, N. (2013). Recommendations for using msBayes to incorporate uncertainty in selecting an ABC model prior: A response to Oaks et al. *Evolution*, 68, 284–294.

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

Loebmann, D., & Haddad, C. F. B. (2010). Amphibians and reptiles from a highly diverse area of the Caatinga domain: Composition and conservation implications. *Biota Neotropica*, 10, 227–256.

Macey, J. R., Strasburg, J. L., Brisson, J. A., et al. (2001). Molecular phylogenetics of western North American frogs of the *Rana boylii* species group. *Molecular Phylogenetics and Evolution*, 19, 131–143.

Magalhaes, I. L. F., Oliveira, U., Santos, F. R., et al. (2014). Strong spatial structure, Pliocene diversification and cryptic diversity in the Neotropical dry forest spider *Sicarius cariri*. *Molecular Ecology*, 23, 5323–5336.

de Melo, W. A., Lima-Ribeiro, M. S., Terribile, L. C., & Collevatti, R. G. (2016). Coalescent simulation and paleodistribution modeling for *Tabebuia rosealba* do not support South American dry forest refugia hypothesis. *PLoS ONE*, 11, e0159314.

Melo Santos, A. M., Cavalcanti, D. R., da Silva, J. M. C., & Tabarelli, M. (2007). Biogeographical relationships among tropical forests in northeastern Brazil. *Journal of Biogeography*, 34, 437–446.

Menezes, L., Canedo, C., Batalha-Filho, H., et al. (2016). Multilocus phylogeography of the treefrog *Scinax eurydice* (Anura, Hylidae) reveals a Plio-Pleistocene diversification in the Atlantic Forest. *PLoS ONE*, 11, e0154626.

Mesquita, D. O., & Colli, G. R. (2003). Geographical variation in ecology of populations of some Brazilian species of *Cnemidophorus*. *Copeia*, 2, 285–298.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Friedrich, L. (2015). *e1071: Misc functions of the Department of Statistics, probability theory group (Formerly: E1071)*. Vienna, Austria: TU Wien.

Monteiro, F. A., Donnelly, M. J., Beard, C. B., & Costa, J. (2004). Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in Northeast Brazil. *Molecular Phylogenetics and Evolution*, 32, 46–56.

Navas, C. A., & Carvalho, J. E. (Eds.) (2010). *Aestivation: Molecular and physiological aspects*. Berlin, Germany: Springer-Verlag.

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97, 145–163.

Nomura, F., Rossa-Feres, D. C., & Langeani, F. (2009). Burrowing behavior of *Dermatonotus muelleri* (Anura, Microhylidae) with reference to the origin of the burrowing behavior of Anura. *Journal of Ethology*, 27, 195–201.

Oaks, J., Sukumaran, J., Esselstyn, J. A., Linkem, C. W., Siler, C. D., Holder, M. T., & Brown, R. M. (2012). Evidence for climate-driven diversification? A caution for interpreting ABC inferences of simultaneous historical events. *Evolution*, 67, 991–1010.

Oliveira, E. F., Gehara, M., Sao-Pedro, V. A., et al. (2015). Speciation with gene flow in whiptail lizards from a Neotropical xeric biome. *Molecular Ecology*, 24, 5957–5975.

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58, 240–242.

Pennington, R. T., Lavin, M., & Oliveira-Filho, A. (2009). Woody plant diversity, evolution, and ecology in the tropics: Perspectives from Seasonally Dry Tropical Forests. *Annual Review of Ecology, Evolution, and Systematics*, 40, 437–457.

Pennington, T. R., Prado, D. E., & Pendry, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, 27, 261–273.

Pereira Filho, G. A., & Montingelli, G. G. (2011). Check list of snakes from the Brejos de Altitude of Paraíba and Pernambuco, Brazil. *Biota Neotropica*, 11, 145–151.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.

Prado, D. E., & Gibbs, P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, 80, 902–927.

Prates, I., Xue, A. T., Brown, J. L., et al. (2016). Inferring responses to climate dynamics from historical demography in Neotropical forest lizards. *Proceedings of the National Academy of Sciences*, 113, 7978–7985.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19, 2092–2100.

Recoder, R. S., Werneck, F. P., Teixeira, M., et al. (2014). Geographic variation and systematic review of the lizard genus *Vanzosaura* (Squamata, Gymnophthalmidae), with the description of a new species. *Zoological Journal of the Linnean Society*, 171, 206–225.

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology*, 12, 196.

Ribeiro, S. C., Roberto, I. J., Sales, D. L., Ávila, R. W., & de Oliveira Almeida, W. (2012). Amphibians and reptiles from the Araripe bioregion, northeastern Brazil. *Salamandra*, 48, 133–146.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.

Thomé, M. T. C., Sequeira, F., Brusquetti, F., et al. (2016). Recurrent connections between Amazon and Atlantic forests shaped diversity in Caatinga four-eyed frogs. *Journal of Biogeography*, 43, 1045–1056.

Wang, X., Auler, A. S., Edwards, R. L., et al. (2004). Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature*, 432, 740–743.

Wang, X., Auler, A. S., Edwards, R. L., et al. (2007). Millennial-scale precipitation changes in southern Brazil over the past 90,000 years. *Geophysical Research Letters*, 34, 1–5.

Werneck, F. P., Costa, G. C., Colli, G. R., Prado, D. E., & Sites, J. W. Jr (2011). Revisiting the historical distribution of Seasonally Dry Tropical Forests: New insights based on palaeodistribution modelling and palynological evidence. *Global Ecology and Biogeography*, 20, 272–288.

Werneck, F. P., Leite, R. N., Geurgas, S. R., & Rodrigues, M. T. (2015). Biogeographic history and cryptic diversity of saxicolous Tropiduridae lizards endemic to the semiarid Caatinga. *BMC Evolutionary Biology*, 15, 94.

Xue, A. T., & Hickerson, M. J. (2015). The aggregate site frequency spectrum for comparative population genomic inference. *Molecular Ecology*, 24, 6223–6240.

Xue, A. T., & Hickerson, M. J. (2017). Multi-DICE: R package for comparative population genomic inference under hierarchical co-demographic models of independent single-population size changes. *Molecular Ecology Resources*, 00, 1–13. https://doi.org/10.1111/1755-0998.12686

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.