MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 1874–1890 doi: 10.1111/mec.12689

Climate, physiological tolerance and sex-biased dispersal shape genetic structure of Neotropical orchid bees

MARGARITA M. LÓPEZ-URIBE,* KELLY R. ZAMUDIO,† CAROLINA F. CARDOSO‡ and BRYAN N. DANFORTH*

*Department of Entomology, Cornell University, Ithaca, NY 14853, USA, †Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA, ‡Laboratório de Sistemática e Ecologia de Abelhas, Universidade Federal de Minas Gerais, Caixa Postal 486, Belo Horizonte, MG 30123-970, Brazil

Abstract

Understanding the impact of past climatic events on the demographic history of extant species is critical for predicting species' responses to future climate change. Palaeoclimatic instability is a major mechanism of lineage diversification in taxa with low dispersal and small geographical ranges in tropical ecosystems. However, the impact of these climatic events remains questionable for the diversification of species with high levels of gene flow and large geographical distributions. In this study, we investigate the impact of Pleistocene climate change on three Neotropical orchid bee species (*Eulaema bombiformis*, *E. meriana* and *E. cingulata*) with transcontinental distributions and different physiological tolerances. We first generated ecological niche models to identify species-specific climatically stable areas during Pleistocene climatic oscillations. Using a combination of mitochondrial and nuclear markers, we inferred calibrated phylogenies and estimated historical demographic parameters to reconstruct the phylogeographical history of each species. Our results indicate species with narrower physiological tolerance experienced less suitable habitat during glaciations and currently exhibit strong population structure in the mitochondrial genome. However, nuclear markers with low and high mutation rates show lack of association with geography. These results combined with lower migration rate estimates from the mitochondrial than the nuclear genome suggest male-biased dispersal. We conclude that despite large effective population sizes and capacity for long-distance dispersal, climatic instability is an important mechanism of maternal lineage diversification in orchid bees. Thus, these Neotropical pollinators are susceptible to disruption of genetic connectivity in the event of large-scale climatic changes.

Keywords: anonymous single copy nuclear loci, euglossine bees, mitochondrial DNA, nuclear DNA, Palaeomodeling, *Pleistocene refugia*

Received 19 October 2013; revision received 20 January 2014; accepted 3 February 2014

Introduction

Climatic and anthropogenic landscape changes are two major threats to the persistence of biodiversity in the 21st century (Heller & Zavaleta 2009; Potts *et al.* 2010). Climate change during the past 50 years has already altered the geographical distribution, developmental time and reproductive rates of many organisms (Parmesan & Yohe 2003; Bartomeus *et al.* 2011). Concurrently, habitat loss and agricultural intensification have fragmented and degraded natural areas, causing severe animal and plant species declines worldwide (Pimm & Kevan 2000). In the face of this biodiversity crisis (Sala *et al.* 2000; Forister *et al.* 2010), an enduring challenge for biologists is to develop a predictive framework to guide appropriate conservation and management plans (Dawson *et al.* 2011). Investigating microevolutionary responses to past geologic and climatic events can provide insights into species vulnerability and predictions

Correspondence: Margarita M. López-Uribe,
Fax: +1 607 255 0939; E-mail: mml82@cornell.edu

about the biodiversity consequences of future environmental changes.

Maintaining evolutionary processes is especially important in areas with high species diversity, such as tropical regions, that are the result of complex ecological and evolutionary histories (Myers *et al.* 2000). For instance, Neotropical lineages have experienced a dynamic geologic history over the past 25 my, including the progressive uplift of the Andes (~23 – ~4.5 mya), the formation of the Amazon basin (~7 mya) and the closure of the Panama Isthmus (~3.5 mya) (Hoorn *et al.* 2010). These tectonic events during the Neogene contributed to species diversification in Neotropical lineages of frogs (Santos *et al.* 2009), bats (Ditchfield 2000) and birds (Fjeldså 1994) through topographical complexity and repeated vicariance/dispersal episodes. More recently over the past 2.6 mya, the geology of the Neotropics has remained relatively unchanged, but climate has been unstable (Bennett 1990). Repeated periods of cold/dry and warm/wet climate altered forest composition, driving shifts in species distributions and intraspecific lineage diversification (Mulcahy *et al.* 2006; Carnaval *et al.* 2009; Prado *et al.* 2012).

Patterns of responses to climatic fluctuations during Pleistocene glaciations vary between Neotropical terrestrial organisms (Hewitt 2004). Phylogeographical studies in vertebrate species with relatively small effective population sizes, low dispersal capacity and restricted distributions often show deep mitochondrial divergences that are temporally congruent with the Pleisteocene, but spatially idiosyncratic, presumably due to differences in species-specific ecological traits (Carnaval *et al.* 2009; Thomé *et al.* 2010; Prado *et al.* 2012; Kieswetter & Schneider 2013). In contrast, the occurrence of strong phylogeographical breaks driven by climatic events in Neotropical species with large population sizes capacity for long-distance dispersal and widespread distributions remains unclear (Cheviron *et al.* 2005; Solomon *et al.* 2008; Martins *et al.* 2009). Therefore, we do not yet have consensus on expected genetic patterns, and potential diversifying mechanisms, for highly mobile terrestrial organisms in the Neotropics.

Understanding the effects of past climatic changes on species' distributions requires the investigation of their demographic responses to environmental change. For explicit hypothesis testing, it is necessary to understand spatial distributions of species in past and present times (Richards *et al.* 2007). The incorporation of ecological niche modelling in phylogeographical studies provides a methodological framework to identify climatically stable areas (CSAs) that served as potential 'refugia' during Quaternary climatic oscillations (Hugall *et al.* 2002; Carnaval *et al.* 2009). This approach generates species-specific spatial models with demographic predictions that can be validated with molecular data (Knowles *et al.* 2007). Furthermore, niche models provide information about species niche breadth and can be used as proxies for animal physiological tolerance (Bonier *et al.* 2007). Although niche conservatism, ecological competition, spatial variation of the environment and dispersal may be the limiting factors to species ranges, geographical distributions are largely determined by climate (Chown *et al.* 2010). Therefore, correlational approaches between species presence data and climate data are used to infer the limits of physiological tolerance for species distributions.

In this study, we investigate the impact of Pleistocene climatic instability on pollinating insects with great dispersal capacities and large population sizes, but varying levels of physiological tolerance to dry climatic conditions. Orchid bees (Apidae: Euglossini) are fast-flying insects endemic to the New World that exhibit an interesting courtship behaviour. Male orchid bees collect volatile compounds from floral and nonfloral resources throughout their lives and present a chemical bouquet to females at display sites enabling mate recognition and choice (Eltz *et al.* 2005). Ecological and behavioural data indicate that dispersal is likely male-mediated in orchid bees because males have large home ranges where they search for volatiles (Wikelski *et al.* 2010). In contrast, females are central-place foragers that exhibit philopatric behaviour (Augusto & Garófalo 2011). Orchid bees are remarkably abundant in low- or mid-elevation forests where they comprise up to 25% of bee communities (Roubik & Hanson 2004). Many orchid bee species have a strong association with wet forested areas and show the highest community richness in forests where precipitation exceeds 2000 mm/year (Dressler 1982). Thus, orchid bees are an excellent system for investigating species responses to the dry–wet Pleistocene climatic cycles in tropical pollinating insects.

We examine the demographic history of three widespread orchid bees of the genus *Eulaema* (*E. bombiformis*, *E. meriana* and *E. cingulata*) using an integrative approach that includes ecological niche modelling and multilocus population genetic data. Specifically, we test whether Pleistocene climate instability led to intraspecific diversification in three orchid bee species with different physiological tolerances. The three focal species of this study have similar widespread distributions, but they display different habitat associations and elevation ranges. Both *E. bombiformis* and *E. meriana* are exclusively found in wet forests, but *E. bombiformis* is distributed from sea level up to 950 m.a.s.l., whereas *E. meriana* is found up to 1700 m.a.s.l. (Ramírez *et al.* 2002). *Eulaema cingulata* is present in both wet and dry lowland forests and its altitudinal distribution reaches 2600 m.a.s.l. (Ramírez *et al.* 2002). Therefore, we predict

these species display a continuum of physiological tolerance to climate variation, with lowest in *E. bombiformis* and *E. meriana* and highest in *E. cingulata*. To test this prediction, we developed a novel method that uses response curves to environmental variables to quantify the relative difference in niche breadth between species.

Based on the shared ecological and behavioural traits, yet differing physiological tolerances among our three focal species, we have three predictions about their responses to historical climatic instability. First, we predict that the amount of suitable habitat in refugia during dry periods of the Pleistocene was reduced for all species, however, not equally. Specifically, *E. cingulata*, which unlike the other two species now inhabits drier forests, should have experienced the largest amount of suitable habitat during the Pleistocene, as wet forests became drier. Second, we predict that species with low physiological tolerance should show strongest phylogeographical structure spatially congruent with hypothesized CSAs because lower physiological tolerance restricted gene flow outside the suitable areas. Third, we predict that the spatial patterns of genetic diversity in these species are the result of isolation/colonization events during repeated cycles of forest contraction/expansion, thus are temporally congruent with Pleistocene climatic instability. The alternative scenario is that these species were widespread throughout the Neotropics pre-Pleistocene and experienced vicariance during the geologic changes in the Pliocene and contraction/expansions events during the Pleistocene. In that case, the deepest phylogeographical breaks in species phylogenies should be temporally congruent with Pliocene geologic events. Testing these predictions has important implications for species persistence under scenarios of future climate change. We discuss the need for further investigation of these questions in species with capacity for long-distance dispersal and the value of integrative studies for future conservation plans of orchid bee species.

Materials and methods

Ecological niche modelling

We identified CSAs for the three *Eulaema* species by overlaying distribution models based on current (averaged over the past 50 years) and last glacial maxima (LGM; 21 kya) climate data through ecological niche models (ENM). We used the current and LGM ENMs as proxies for species-specific distributions during interglacial and glacial periods, respectively. We compiled current presence data points for the three focal species from published articles, the Discover Life Website, the Global Biodiversity Information Facility, and data points from individuals collected during our own fieldwork. In total, we collected 101, 127 and 123 location data points for *E. bombiformis*, *E. meriana* and *E. cingulata*, respectively (Fig. S1, Supporting information). Species distributions were modelled in MAXENT v. 3.3.3 (Phillips *et al.* 2006; Phillips & Dudík 2008) using 19 environmental data layers at a 30-arc-second (~1 km) resolution from the WORLDCLIM database. All bioclimatic variables were included in analyses despite correlation between them because the exclusion of correlated variables usually has little effect on ENMs (Knowles & Alvarado-Serrano 2010). We evaluated the predictive power of the model in two ways. First, we used a binomial test based on omission and predicted areas to test whether our model predicted species distributions better than random. Second, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to measure model performance based on the sensitivity and specificity of the model. A random sample of 25% of the initial data set over 10 replicates was used as 'testing data' to check for the spatial independence of the locality points used to build the models. We collected data from ten independent runs to verify the replicability of our models. To define binary presence–absence suitability models, we selected the MAXENT threshold that minimized the difference between the omission rate and predicted area (equal training sensitivity and specificity threshold) (Pearson *et al.* 2004).

Assuming species niches have not changed considerably during the past 21 000, we characterized species distributions during LGM by projecting results from the current distribution models onto the same environmental layers from the LGM. We overlaid the binary predicted distributions of the current and LGM models using the raster calculator tool in ARCGIS 10 (ESRI 2011) to outline areas that have been climatically stable throughout the repeated glacial cycles of the past 1.5 My. We defined major 'refugia' as contiguous areas with a minimum size of 2000 km^2 and separated by a minimum linear distance of 300 km, a distance 10 times greater than the maximum home range distance reported for orchid bees (Janzen 1971).

Niche breadth quantification

We used response curves to all bioclimatic variables to calculate the relative ecological niche breadth for our focal species. Response curves depict the range of species' predicted suitable conditions in ecological space based on individual responses to one bioclimatic variable (Phillips & Dudík 2008). Response curves can provide information on what niche space a species can tolerate, but they are uninformative about what conditions species cannot tolerate because ranges may also be restricted by biotic factors (Chown *et al.* 2010).

Therefore, we developed the relative niche breadth (RNB) index to serve as a proxy for the species-specific breadth of physiological tolerance and quantify the relative difference in tolerance between species. We calculated this index as follows:

$$\mathrm{RNB} = \frac{\sum_j^m \Delta_{ij}}{\max\left\{\sum_j^m \Delta_{ij}\right\}_{i=1}^{j=n}} \quad (1)$$

where Δ_{ij} is the difference between the minimum and the maximum value of the environmental variable j in species i for which the probability of presence is at least 0.5 and $\max\left\{\sum_j^m \Delta_{ij}\right\}_{i=1}^{j=n}$ is the maximum sum of all Δ_{ij} for all environmental variables (m) among all species (n) compared (Fig. S2, Supporting information). The rationale for this index as a proxy for physiological tolerance breadth is that the greater the difference between the minimum and maximum probability of species' presence for a given environmental variable, the wider the range of environmental conditions that a species can tolerate. The 0.5 probability value is used as a benchmark for the delta calculation because it is a standard cut-off threshold for the probability of presence of a species (Trevan 1927). This cut-off also avoids the potential ambiguity of making measurements from the extreme of the curves where values tend to converge.

Sampling

We collected a total of 454 male individuals from throughout the geographical distribution of *E. bombiformis* (113), *E. meriana* (201) and *E. cingulata* (137) from 14, 16 and 16 localities, respectively (Fig. 1). In the field, we attracted male orchid bees using five chemical baits (cineol, methyl salicylate, vanillin, eugenol and benzyl acetate) from several sites at each locality to avoid sampling-related individuals. Male bees were collected using entomological nets and preserved in 95% ethanol. Geographical positioning was recorded for each locality using a Garmin GPSmap 76CSx unit. Front legs from all bees were removed for DNA extraction before individuals were mounted on entomological pins, labelled and accessioned in the Cornell University Insect Collection (CUIC). Field collections were supplemented with museum specimens from localities in Guatemala, Bolivia, Colombia, French Guiana and Brazil (Table S1, Supporting information).

Genetic data collection

Leg tissue was ground in liquid nitrogen and digested overnight in 1 μg/μL of Proteinase K. DNA was extracted using the DNeasy blood and tissue extraction

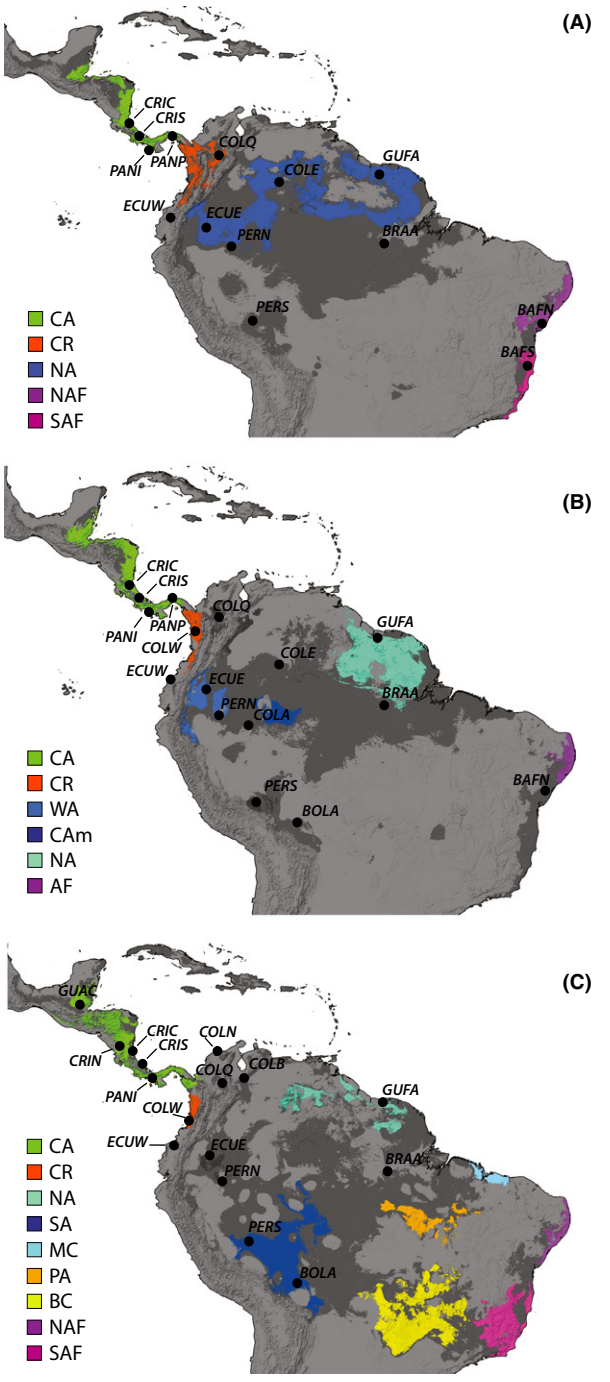

Fig. 1 Climatically stable areas (CSA) and distribution of sampling localities for *Eulaema bombiformis* (A), *E. meriana* (B) and *E. cingulata* (C). CSA are the result of overlapping current and LGM climatic suitability models for each species. Colored areas indicate contiguous CSA: Central America (CA), Chocó Region (CR), Northern Amazon (NA), Northern Atlantic Forest (NAF), Southern Atlantic Forest (SAF), West Amazon (WA), Central Amazon (CAm), Southern Amazon (SA), Brazilian Cerrado (BC), Pará Amazon (PA) and Maranhão Coast (MC). Shaded grey areas indicate suitable areas for the three species under current climatic conditions.

kit (QIAGEN) for field collected tissues, or a standard phenol–chloroform DNA extraction protocol in the case of museum specimens (Danforth 1999). We sequenced a total of ~3750 base pairs from seven genes: two mitochondrial genes (COI: Cytochrome oxidase I, and Cyt*b*: Cytochrome *b*), two nuclear protein-coding genes (CAD: carbamoyl-phosphate synthetase 2, aspartate transcarbamylase and dihydroorotase; and EF1-α: Elongation Factor 1 alpha) and three anonymous single copy nuclear loci (ASCN) that included a microsatellite region (EM8, EM70 and EM106) (Table 1). We chose these markers because they all showed clear sequencing results and comprise a group of genes with a wide range of mutation rates. All mitochondrial and nuclear coding genes were sequenced in both directions, while the ASCN loci were sequenced only in the forward strand using a 24-bp tag added to the forward primers (López-Uribe *et al.* 2011). Fragments were sequenced using the Big Dye terminator kit (version 3.1) (Applied Biosystems) on an ABI 3730 capillary sequencer.

Genetic diversity

Haploid male sequences were assembled and individually edited in SEQUENCHER v.5.0 (DNASTAR). Haplotype sequences were deposited in GenBank (Accession nos KF895558–KF970454; Table S1, Supporting information). We used Clustal W to generate a multiple sequence alignment in MEGALIGN v.8.0.2 (DNASTAR), which was then manually edited in MACCLADE v.4.08 (Maddison & Maddison 2005). We calculated the number of polymorphic sites (N_P), number of singletons (N_S) and number of alleles (N_A) of each gene per species in the software DNASP v.5 (Librado & Rozas 2009). We also estimated the average number of nucleotide differences (k), nucleotide diversity (π) and haplotype diversity (h) of the populations grouped by the CSAs identified based on ENMs. Because mitochondrial recombination is possible (Galtier *et al.* 2009), we tested for recombination in each protein-coding gene and in the concatenated mitochondrial data set using Sawyer's statistical test, maximum chi-squared, sister scanning method and automated bootscanning as implemented in the software RDP (Martin *et al.* 2005).

Phylogenetic inference

We inferred the best-fit models of molecular evolution for the concatenated data set of protein-coding genes using JMODELTEST v.0.1.1 (Posada 2008) based on the corrected Akaike information criterion (AICc). Bayesian analyses were performed separately for each species in MRBAYES v.3.1.2 (Huelsenbeck & Ronquist 2001) using two independent runs each with three heated and one cold Markov chains sampling every 1000 generations. Twenty million generations were sufficient to ensure the average standard deviation of split frequencies was

Table 1 Summary of PCR primers and genetic diversity indices per locus for *Eulaema bombiformis* (Ebom), *E. meriana* (Emer) and *E. cingulata* (Ecin)

Locus	Primers	Species	Length	N_P	N_S	N_A	Reference
COI	F: 5′-CAA CAT TTA TTT TGA TTT TTT GG-3′	Ebom	838	51	15	38	Simon *et al.* (1994)
	R: 5′-TCC AAT GCA CTA ATC TGC CAT ATT A-3′	Emer	847	53	12	59	
		Ecin	835–838	23	9	25	
Cytb	F: 5′-TAT GTA CTA CCA TGA GGA CAA ATA TC-3′	Ebom	433	18	4	24	Crozier *et al.* (1991)
	R: 5′-ATT ACA CCT CCT AAT TTA TTA GGA AT-3′	Emer	433	39	11	45	
		Ecin	433	18	5	26	
CAD	F: 5′-GGW TAT CCC GTD ATG GCB MGW GC-3′	Ebom	673	40	16	63	Danforth *et al.* (2006)
	R: 5′-GCA THA CYT CHC CCA CRC TYT TC-3′	Emer	673	17	6	69	
		Ecin	673	46	30	69	
EF1α	F: 5′-G GGY AAA GGW TCC TTC AAR TAT GC-3′	Ebom	991–994	33	26	10	Danforth & Ji (1998)
	R: 5′-A ATC AGC AGC ACC TTT AGG TGG -3′	Emer	991–996	17	14	10	
		Ecin	991–994	29	25	13	
EM8	F: 5′-CAG CGT CGC GAT TGG TTC TAC A-3′	Ebom	210–238	5	5	6	López-Uribe *et al.* (2011)
	R: 5′-TCA GCT TTG TCA CCG GCA CTG T-3′	Emer	260–282	12	3	13	
		Ecin	254–268	10	6	9	
EM70	F: 5′-GTA CCA CTG CGA GAG CGA AGA AAA-3′	Ebom	229–240	13	9	12	López-Uribe *et al.* (2011)
	R: 5′-CCA GTG GCC CGA AGT AGA AAC A-3′	Emer	231–254	15	8	11	
		Ecin	231–251	27	15	33	
EM106	F: 5′-GAC GTG GAT GAG CCG CAG AAG AC-3′	Ebom	207–234	15	1	24	López-Uribe *et al.* (2011)
	R: 5′-TCC GAC GAT GTA CGA GCA CGA A-3′	Emer	197–257	6	1	11	
		Ecin	219–267	12	7	12	

N_P, Number of polymorphic sites; N_S, number of singletons; N_A, number of alleles.

below 0.01. Stationarity of likelihood scores and parameter convergence was assessed using TRACER v. 1.4. Posterior probabilities were estimated from the distribution of trees after discarding the first 20 000 as burn-in. We used sequences from the other two *Eulaema* spp. included in this study as outgroups for phylogenetic reconstructions.

A time-calibrated phylogeny of the mitochondrial data set was estimated in BEAST 1.6.2 (Drummond & Rambaut 2007) under the coalescent model of expansion growth. We estimated the time to the most recent common ancestor (tMRCA) using the COI substitution rate of 1.2–1.5% per million years calibrated for other insects (Farrell 2001). Node ages were estimated using a strict clock, which is appropriate for intraspecific data analysed under a coalescent model. We performed two independent runs for 20 million generations sampling every 1000 generations. Convergence was assessed by examining run log files in TRACER v. 1.4 and confirming that effective sampling size (ESS) for parameters was greater than 200. To reconstruct the maximum clade credibility (MCC) tree, we discarded the first 10% of the sampled trees as burn-in.

Haplotype networks

Mitochondrial and nuclear protein-coding haplotypes were also visualized with a neighbour-net network constructed in SPLITSTREE4 v. 4.12.3 (Huson & Bryant 2006). We used the Tamura–Nei distance to estimate the net sequence divergence between major phylogeographical groups in the program DNASP v.5 (Librado & Rozas 2009). To detect spatial structuring in the mitochondrial vs. nuclear data, we used the global and local tests of the spatial principal component analysis (sPCA) in the R package ADAGENET v.1.3-9.2 (Jombart *et al.* 2008; R Core Team 2013). sPCA uses allele frequencies to investigate patterns of genetic variability in a spatially explicit context. This test differentiates between global structures that indicate the presence of coarse-scale spatial patterns (e.g. clines, patches), and local structures, that detect differentiation between neighbouring sites. This approach is appropriate for detecting spatial patterns unidentifiable with tree-based methods (Jombart *et al.* 2008).

Population structure

We tested phylogeographical hypotheses based on the CSAs identified for each species (Table S2, Supporting information) using two different methods. In all cases, mitochondrial sequence data and nuclear haplotype data were analysed separately due to potential for incongruent phylogeographical signals from the different markers. First, we used full and partial Mantel tests, as implemented in IBDWB v.3.23 (Jensen *et al.* 2005), to investigate patterns of isolation by distance (linear geographical distance) and isolation-by-barrier (matrix between hypothesized CSA) as predictors of the level of genetic differentiation between populations. Interpopulation differentiation was tested using Rousset's distance method [PhiST/(1-PhiST)]. Second, we implemented an analysis of molecular variance (AMOVA) in ARLEQUIN v. 3.5.1.2 to calculate the percentage of genetic variation explained by populations grouped based on the identified CSAs (Excoffier & Lischer 2010).

Demographic inference

We calculated Fu's *Fs* in ARLEQUIN v. 3.5.1.2 (Excoffier & Lischer 2010) to detect departures from neutrality or constant population size. The demographic history of each phylogeographical lineage was reconstructed using an Extended Bayesian Skyline Plot (EBSP) as implemented in BEAST 1.6.2 (Drummond & Rambaut 2007). All nuclear and mitochondrial markers were pooled in this analysis, and the haplodiploid effective population size of nuclear markers was incorporated in our model as x-linked genes. To test for sex-biased dispersal, we estimated migration rates between CSAs for nuclear and mitochondrial data independently by fitting a model of isolation-with-migration as implemented in the software IMa2 (Hey & Nielsen 2007). Greater migration rates in the nuclear than in the mitochondrial DNA would support the male-biased dispersal hypothesis. For loci EM8 and EM70, information about the microsatellite and flanking regions was provided separately. Locus EM106 was not included because the flanking region violated the infinite sites mutation model. First, we analyzed simultaneously all populations of *E. bombiformis* and *E. meriana*. However, due to the computational challenges of estimating a high number of parameters from multiple populations, the IMa2 analyses only reached convergence for pairs of sister populations. Thus, we only present results for two pairs of populations from neighbouring CSAs in *E. bombiformis* (pair 1: 'CR2' and CA; pair 2: NA2 + NA3 and 'AF'). Final analyses consisted of two runs of 10–40 geometrically heated chains with burn-in 100k steps and sampled for 20 000 000 steps. We report population migration rate calculated as the mutation-scaled migration rate multiplied by population size.

Results

Niche modelling

The MAXENT ENMs accurately predicted the current distribution of the three species (AUCEbom = 0.983 ± 0.001;

AUCEmer = 0.989 ± 0.002; AUCEcin = 0.968 ± 0.013) (Fig S1 A–C, Supporting information). However, we identified some areas of under-prediction likely due to limited locality presence records from these regions (Fig S1 A–C, Supporting information). Performance of the test and training data sets were similar, indicating that the data points used for model building were not spatially autocorrelated. The permutation importance test indicated that temperature seasonality was the variable contributing the most to model gain in all three species. Precipitation of the coldest quarter was also an important variable for *E. bombiformis* and *E. meriana*.

For *E. bombiformis*, five major CSAs were identified (Fig. 1A) with two major distribution shifts during LGM: (i) a reduction in the suitable habitat in the Amazon basin to one large suitable area located on the Northern Amazon and the foothills of Guiana highlands (NA) and (ii) a subdivision of the Atlantic Forest into two smaller refugia (NAF, SAF) spatially congruent with earlier models of Pleistocene refugia (Carnaval & Bates 2007; Carnaval *et al.* 2009). The distribution of *E. meriana* was more fragmented during LGM, and six major CSAs were identified (Fig. 1B). For *E. meriana*, the suitable habitat in the Atlantic Forest was reduced to one small area restricted to the coast of Pernambuco (AF). The Amazon Basin was reduced to small pockets of forest located to the west and centre of the Amazon (WA, CAm) and the foothills of Guiana highlands (GF). For *E. cingulata*, we identified eight CSAs with the largest areas located in Central America and the southern part of the current distribution of this species (Fig. 1C). In contrast to *E. bombiformis* and *E. meriana*, CSAs for *E. cingulata* included current dry forests and savanna areas, such as the Brazilian Cerrado (BC).

Niche breadth index

Relative niche breadth indices showed that the species with the broadest niche breadth is *E. cingulata* (RNB = 1; $\Sigma_{\Delta Ecin}$ = 54 131) followed by *E. bombiformis* (RNB = 0.34; $\Sigma_{\Delta Ebom}$ = 18 396) and *E. meriana* (RNB = 0.33; $\Sigma_{\Delta Emer}$ = 17 748) (Table S3, Supporting information). More specifically, *E. meriana* and *E. bombiformis* showed one-third the niche breadth of *E. cingulata*. This difference is in accordance with the exclusive association of *E. meriana* and *E. bombiformis* with wet forests, while *E. cingulata* inhabits both wet and dry forests. Response curves to environmental variables revealed that *E. bombiformis* has broader physiological tolerance than *E. meriana* based on 16 of the 19 climatic variables. However, *E. meriana* has a wider physiological tolerance to colder temperatures (ΔBIO6 – Table S3, Supporting information), suggesting that this physiological trait may explain its greater altitudinal range. Overall, palaeomodels indicate that suitable habitat was reduced during the dry period of the Pleistocene for all three species, especially in the Amazon Basin (Fig. S1D–F, Supporting information). As expected, the species with the broadest physiological tolerance (*E. cingulata*) had the most widespread distribution during dry climatic periods of the Pleistocene despite experiencing the most severe reduction in overall suitable habitat (47%). Reduction in suitable habitat was 30% and 40% for *E. bombiformis* and *E. meriana*, respectively.

Patterns of genetic diversity

The degree of mitochondrial genetic diversity within CSAs was concordant with predictions based on the age of each phylogeographical group (see below). Generally, geographical areas with older lineages showed greater levels of genetic diversity. Geographical regions that showed the highest mitochondrial genetic diversity were the Northern Amazon (NA) in *E. bombiformis* ($\bar{\pi}_{mt}$ = 0.00751), Chocó Region (CR) in *E. meriana* (π_{mt} = 0.00982) and Northern Amazon (NA) in *E. cingulata* ($\bar{\pi}_{mt}$ = 0.00472) (Table 2). Average nucleotide diversity was higher for mitochondrial than for nuclear genes in *E. meriana* ($\bar{\pi}_{mt}$ = 0.00548; $\bar{\pi}_{nu}$ = 0.00286), but similar in *E. bombiformis* ($\bar{\pi}_{mt}$ = 0.00454; $\bar{\pi}_{nu}$ = 0.00423) and *E. cingulata* ($\bar{\pi}_{mt}$ = 0.00337; $\bar{\pi}_{nu}$ = 0.00312). We found no evidence of recombination for the mitochondrial or nuclear genes using four different recombination detection tests.

Time-calibrated phylogenies

Bayesian tree topologies of concatenated mtDNA phylogenies were congruent between the MRBAYES and BEAST analyses. Mitochondrial Bayesian phylogenies recover strong phylogeographical structure for *E. bombiformis* and *E. meriana*, the two focal species with narrower physiological tolerances, but not for *E. cingulata* (Fig. 2). Furthermore, the majority of the lineages in *E. bombiformis* and *E. meriana* were spatially congruent with predicted CSAs and date to the Pleistocene (Fig. 2). For *E. bombiformis*, we analysed both mitochondrial genes concatenated under the HKY + I + G model. The MCC tree recovered one old lineage from the Amazon basin (Am1) and two reciprocally monophyletic groups [~1.5 Mya; 95% highest posterior density (HPD) = 0.88, 2.16] spatially concordant with an east–west split due to the presence of the Andes cordillera (Fig. 2A). Within the east clade, we found three monophyletic lineages from the Amazon Basin (Am1, Am2 and Am3) and one lineage that groups all the Atlantic Forest haplotypes and two individuals from the Amazon basin ('AF') (Fig. 2A). Within the west clade, samples from Central

Table 2 Genetic diversity indices for the mitochondrial (mtDNA) and nuclear (nuDNA) protein-coding genes of populations grouped by the closest CSA in *E. bombiformis* (Ebom), *E. meriana* (Emer) and *E. cingulata* (Ecin). Average nucleotide differences (*k*), nucleotide diversity (π) and haplotype diversity (*h*)

		mtDNA				nuDNA			
Species	CSA	*n*	*k*	π	*h*	*n*	*k*	π	*h*
Eulaema bombiformis	CA	19	3.520	0.00277	0.708	10	7.667	0.00330	1
	CR	10	6.578	0.00518	0.956	8	13.071	0.00563	1
	NA	24	9.543	0.00751	0.989	18	10.261	0.00444	1
	AF*	10	3.444	0.00271	0.911	9	8.194	0.00354	1
Eulaema meriana	CA	67	6.316	0.00493	0.95	49	6.276	0.002669	0.999
	CR	31	12.568	0.00982	0.916	29	7.320	0.003082	1
	Am†	31	3.596	0.00290	0.94	38	6.434	0.002715	1
	GF	5	6.2	0.00484	1	6	6.667	0.002805	1
	AF	7	6.286	0.00491	0.857	4	7.167	0.003011	1
Eulaema cingulata	CA	62	4.303	0.00339	0.887	44	8.309	0.00351	1
	NA	9	5.99	0.00472	0.989	4	7.5	0.00314	1
	WA	25	3.94	0.00311	0.903	16	7.283	0.00306	0.992
	SA	12	2.879	0.00227	0.955	9	6.556	0.00275	1

*AF groups populations from NAF and SAF.
†Am groups populations from WA and CAm.

America (CA) are monophyletic (~0.2 Mya; HPD = 0.09, 0.39), while individuals from the Chocó Region ('CR1', 'CR2') form two highly supported clades that include some individuals from Central America.

The MCC tree for *E. meriana*, based on the GTR + I + G model, revealed two old lineages that diverged early in the history of this species: one from Central America (CA1) (~4.3 Mya; HPD = 1.96, 5.80) and one from Chocó region (CR1) (~2.9 Mya; HPD = 1.61, 4.36) (Fig. 2B). These early-diverged lineages are sister to a clade (~1.6 Mya; HPD = 0.93, 2.43) that comprises two lineages: one from the Amazon Basin (Am) (~0.44 Mya; HPD = 0.17, 0.75) + Atlantic Forest ('AF') (~0.96 Mya; HPD = 0.41, 1.57) sister to a monophyletic clade comprising individuals from Guiana's foothills (GF), Chocó Region ('CR2') (~0.42 Mya; HPD = 0.20, 0.67) and Central America ('CA2') (~0.33 Mya; HPD = 0.19, 0.51).

The Bayesian phylogenetic trees for *E. cingulata* were built using the HKY + I molecular evolution model. Unlike the patterns found in *E. bombiformis* and *E. meriana*, the mitochondrial data set recovered no geographically structured lineages in *E. cingulata* (Fig. 2C). The only obvious features of the tree topology are the presence of (i) only one highly supported phylogenetic group with mixed haplotypes from Central America, Chocó Region and Amazon basin and (ii) a monophyletic lineage that includes haplotypes from the Southern Amazon basin refugia (Am).

The Bayesian phylogenetic reconstruction from the protein-coding nuclear genes did not recover phylogeographical signal for any of the species (not shown). Identical nuclear haplotypes were found in individuals from Costa Rica and Bolivia located ~4000 km apart. Although not informative for phylogenetic inference, the haplotype genetic data from the nuclear loci were used for the population and phylogeographical level analyses (see below).

Average mitochondrial nucleotide divergence between phylogeographical groups was 1.13% for *E. bombiformis*, 0.96% for *E. meriana* and 0.39% for *E. cingulata*. Average nuclear nucleotide divergence was lower than mitochondrial divergence in *E. bombiformis* and *E. meriana* (0.53% and 0.3%, respectively) but similar in *E. cingulata* (0.31%).

Haplotype networks and microsatellite trees

Neighbour-net haplotype networks for the mitochondrial data strongly supported the spatial grouping based on the ENMs for *E. bombiformis* and *E. meriana* (Fig. 3A,B). The sPCA global test indicated highly significant structure for all species in the mitochondrial data set ($p_{Ebom} < 0.001$; $p_{Emer} < 0.001$; $p_{Ecin} < 0.001$) (Fig. S3, Supporting information). These results reveal cryptic genetic structure that is not explained by the hypothesized CSAs in *E. cingulata*. The nuclear haplotype network data for CAD and EF1α showed a random distribution of haplotypes based on their geographical origin (Fig. 3D–I). These results are supported by the nonsignificant sPCA global test results for all species ($p_{Ebom} = 0.131$; $p_{Emer} = 0.208$, $p_{Ecin} = 0.347$).

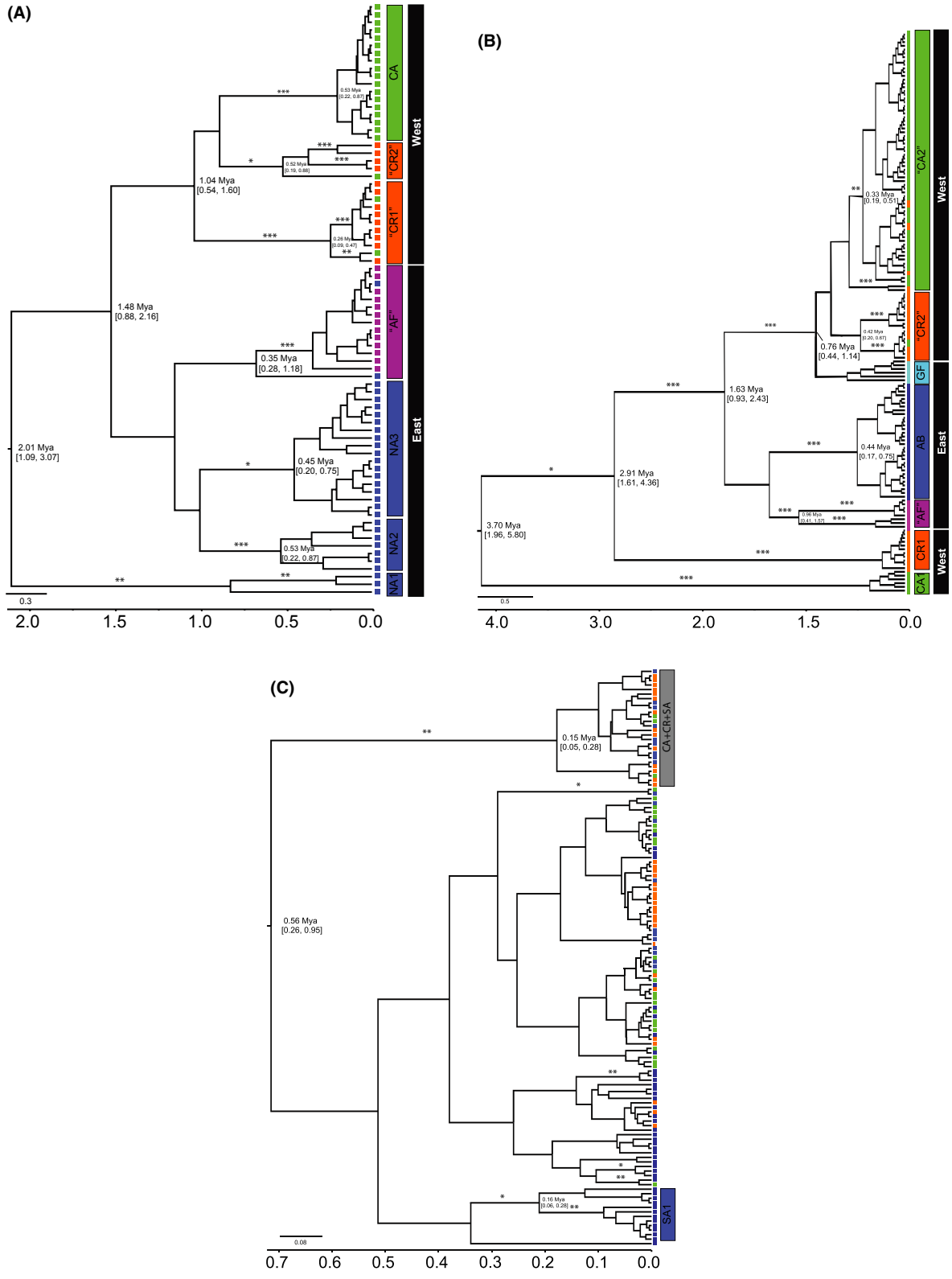

Fig. 2 Calibrated mitochondrial phylogenies for *E. bombiformis* (A), *E. meriana* (B) and *E. cingulata* (C). Asterisks represent Bayesian posterior probability branch supports (*>70%, **>80%, ***>90%). Bayesian estimates for the time to the most recent common ancestor (rMRCA) in million years (my) of crown groups are shown in front of each node. Colours correspond to the climatically stable areas (CSAs) identified from the ecological niche models (Fig. 1). See main text for details.

Local structure was detected only for *E. meriana* in both the mitochondrial and nuclear datasets (mt: $p_{Ebom} = 1$; $p_{Emer} = 0.026$, $p_{Ecin} = 0.097$; nu: $p_{Ebom} = 0.12$; $p_{Emer} = 0.007$, $p_{Ecin} = 0.671$) (Fig. S3, Supporting information). This significant local structure is likely the result of genetic differentiation between *E. meriana* individuals codistributed in Central America and the Chocó Region but belonging to the basal and derived lineages in this species (Fig. 2B).

Phylogeographical structure and signatures of population expansion

The hierarchical *F*-statistics results for the mitochondrial and nuclear data sets were incongruent across all three focal species. The AMOVA analysis of mtDNA data indicated that CSAs explained 24% and 34% of the variation in *E. bombiformis* and *E. meriana*, respectively, but did not explain any variation in *E. cingulata* (Table 3). Likewise, Fct values based on mtDNA were highly significant in *E. bombiformis* and *E. meriana*, but not in *E. cingulata*. For nuDNA, most of the genetic variation was found within populations (Vc > 90% for all three species) (Table 3). We found significant patterns of isolation by distance for the mitochondrial data in *E. bombiformis* and *E. meriana* ($r_{Ebom} = 0.32$, $P = 0.031$; $r_{Emer} = 0.52$, $P = 0.001$), but not in *E. cingulata* ($r_{Ecin} = 0.0082$, $P = 0.28$). However, isolation-by-barrier showed a higher correlation with genetic differentiation in *E. bombiformis* and *E. meriana* ($r_{Ebom} = 0.65$, $P = 0.001$; $r_{Emer} = 0.67$, $P < 0.001$). Neither IBD nor IBB were significant for the nuclear haplotype data (IBD: $r_{Ebom} = 0.19$, $P = 0.12$; $r_{Emer} = 0.12$, $P = 0.22$; $r_{Ecin} = 0.16$, $P = 0.14$; IBB: $r_{Ebom} = -0.018$, $P = 0.47$; $r_{Emer} = 0.13$, $P = 0.88$; $r_{Ecin} = 0.0078$, $P = 0.53$).

Fu's F_s values were negative and highly significant in all phylogeographical groups except for NA and AF in *E. meriana* (Table 4). These results indicate that most of the sampled populations have not reached mutation–drift equilibrium due to demographic population expansion or purifying selection. The EBSPs detected one population size increase for all lineages except for CR in *E. bombiformis* and AF in *E. meriana* (Fig. 4). EBSPs were not calculated in *E. cingulata* due to lack of monophyletic lineages in this taxon. Estimates of current effective population sizes based on EBSPs vary between 66 000 and 37 000 000 individuals per lineage (Table 4). IMa2 runs for two pairs of populations in *E. bombiformis* (pair 1: 'CR2' and CA; pair 2: NA2 + NA3 and 'AF') reached convergence with high ESS values across all parameters. However, parameters could not be estimated with certainty (posterior density did not reach low levels at the lower limit of the prior); thus, these results should be interpreted with caution. Peak posterior distribution of migration rate estimates indicated that the number of migrants per generation was at least three times lower in the mtDNA than in the nuDNA (Table 5, Fig. S4, Supporting information). The highest estimated migration rate was detected from Central America (CA) to the Chocó Region ('CR2') (mtDNA; HiPt 0.0125, 95% HPD = 0.0075–97.49; nuDNA: HiPt 0.563, 95% HPD = 0.168–22 120.40) and the lowest from the Atlantic Forest ('AF') to the Amazon (NA2 + NA3) (mtDNA: HiPt 0.015, 95% HPD = 0.0084–244.76; nuDNA: HiPt 0.619, 95% HPD = 0.219–296.41) (Table 5, Fig. S4, Supporting information).

Discussion

Demographic effects of climatic stability

Ours is the first study to investigate the phylogeographical history of widespread pollinators at a transcontinental geographical scale using a combination of spatial models and multilocus molecular markers. Our results confirmed our hypothesis that climatic instability during the Pleistocene played an important role on the intraspecific lineage diversification in these Neotropical pollinators and that the impact of historical climate variability varied between taxa with different niche breadths. Our initial hypothesis was supported based on three results. First, species-specific palaeomodels indicated a reduction in the geographical distribution of all three species. As predicted, *E. cingulata* was the most widely distributed species during the LGM, but this species also experienced the greatest reduction in distribution, suggesting that during glaciations, the dry areas in the Neotropics were unsuitable even for the species with the widest physiological tolerance. Second, our results revealed strong geographical structuring of mitochondrial genetic diversity in the two species with narrower physiological tolerance, consistent with predictions of population persistence in refugia and low female dispersal between CSAs. This pattern was not found in *E. cingulata*, the species with wide physiological tolerance. The lack of structuring in *E. cingulata* may be the result of high historical and contemporary dispersal between CSAs, given the broad physiological tolerance of this species. Even though the ENM for *E. cingulata* indicated a dramatic reduction in habitat suitability during the Pleistocene, we cannot dismiss the possibility that this species was capable of dispersal through unsuitable areas during that time. Third, phylogeographical structuring was spatially and temporally consistent with the inferred CSAs. Our dating analysis indicated that both *E. bombiformis* and *E. meriana* originated during late Pliocene and, except for the two basal lineages in *E. meriana*, all lineages date from the

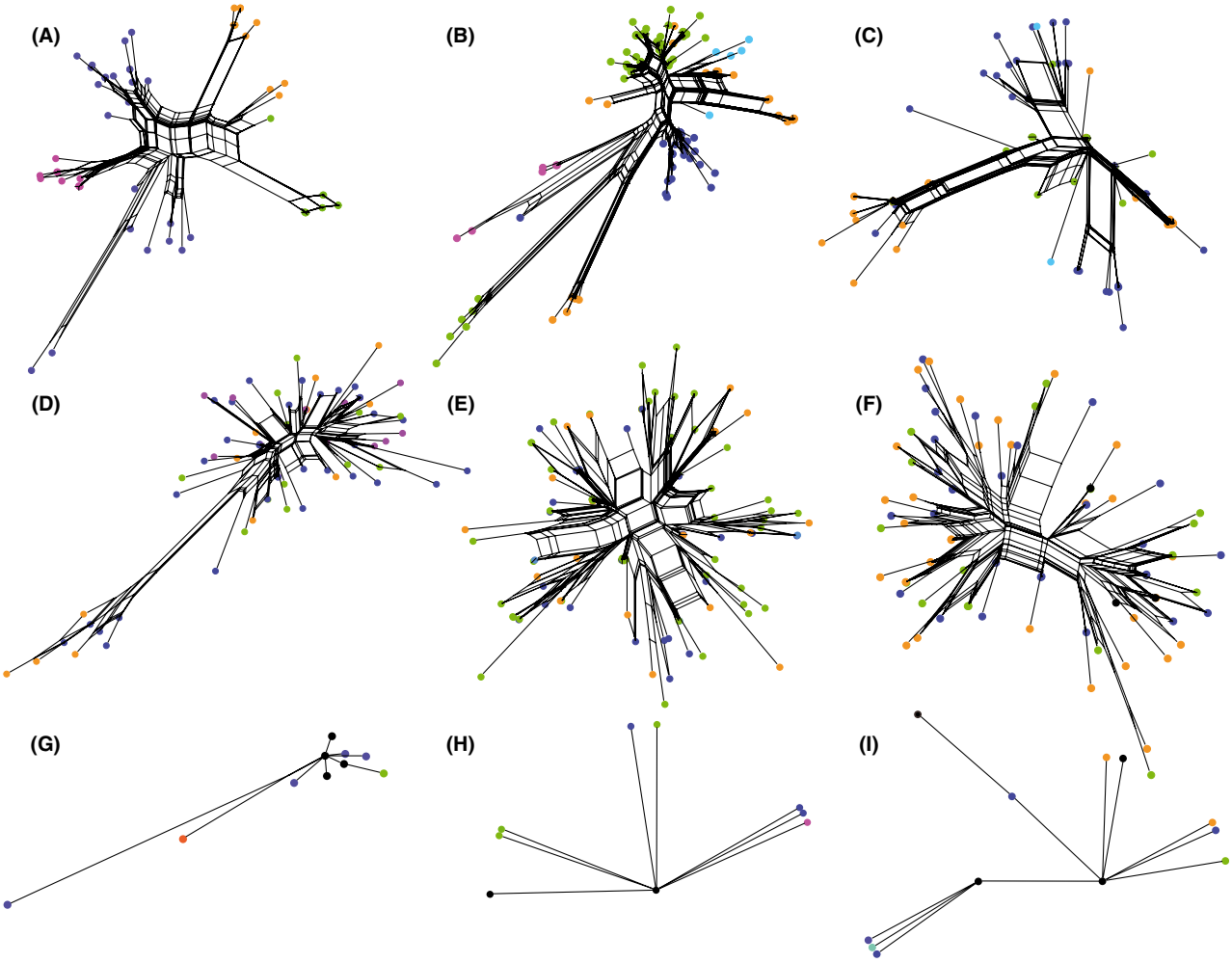

Fig. 3 Multilocus mitochondrial (A–C), CAD (D–F) and EF1α (G–I) networks under the neighbour-net algorithm for *E. bombiformis* (A, D, G), *E. meriana* (B, E, H) and *E. cingulata* (C, F, I). Haplotypes are coloured by the closest CSAs (Fig. 1, Table S2) from their geographical origin. Black dots indicate haplotypes that were found in multiple CSAs. CSAs, Climatically stable areas.

Table 3 Summary of the analysis of molecular variance (AMOVA) for the mitochondrial (mtDNA) and nuclear (nuDNA) data sets. Populations are grouped based on the identified CSAs from the ecological niche modelling

		Among CSA		Among populations within CSA		Within populations	
Species		Va (%)	F_{ct}	Vb (%)	F_{sc}	Vc (%)	F_{st}
Eulaema bombiformis	mtDNA	23.71	0.24**	8.56	$0.11^{n.s}$	67.73	0.32***
	nuDNA	5.23	0.053*	−2.76	$-0.025^{n.s}$	97.53	$0.025^{n.s}$
Eulaema meriana	mtDNA	34.33	0.34***	11.31	0.17***	54.36	0.46***
	nuDNA	2.65	$0.026^{n.s.}$	6.27	0.064**	91.07	0.089***
Eulaema cingulata	mtDNA	−3.13	$-0.031^{n.s.}$	39.95	0.39***	63.18	0.37***
	nuDNA	0.005	$0.002^{n.s.}$	−0.02	$-0.009^{n.s}$	100	$-0.008^{n.s.}$

Significance of *P*-values shown as < 0.05 (*), *P* < 0.01 (**), *P* < 0.005 (***), or nonsignificant (n.s).

Pleistocene, suggesting that the diversification of these lineages is recent, and not likely a result of earlier Pliocene geological and tectonic events in South America.

We found striking incongruence in the patterns of distribution of genetic diversity from nuclear and mitochondrial markers. MtDNA showed a significant association

Table 4 Demographic parameter estimates for lineages and neutrality tests for geographical groups in *E. bombiformis*, *E. meriana* and *E. cingulata*. Current median effective population size (*Ne* in thousands) and number of changes in population size (Δ*Ne*) were estimated through Extended Bayesian Skyline Plots

	Lineage	*Ne*	Δ*Ne*	CSA	Fu's *Fs*
Eulaema	CA	377 (±242)	1 (±0.042)	CA	−21.834***
bombiformis	'CR1'	419 (±291)	0 (±0.001)	CR	−4.846**
	NA2	2049 (±169)	1 (±0.0087)	NA	−6.4571*
	NA3	3714 (±1563)	1 (±0.013)		
	AF	371 (±96.4)	1 (±0.096)	AF	−7.609***
Eulaema	CA1	246 (±5.6)	1 (±0.005)	CA	−25.184**
meriana	'CA2'	670 (±121)	1 (±0.021)		
	CR1	29 (±16.4)	1 (±0.019)	CR	−22.562***
	'CR2'	243 (±198)	1 (±0.033)		
	NA	66 (±16)	1 (±0.009)	NA	−1.0114
	WA	1622 (±558)	1 (±0.018)	WA	−26.035***
	AF	689 (±260)	0 (±0.011)	AF	−2.482
Eulaema	CA	N.A.	N.A.	CA	−26.495***
cingulata	CR	N.A.	N.A.	CR	−25.231***
	NAm	N.A.	N.A.	NAm	−8.627***
	SAm	N.A.	N.A.	SAm	−25.626***

Significance for the neutrality tests shown as $P < 0.02$ (*), $P < 0.005$ (**), $P < 0.001$ (***), or nonsignificant (n.s).

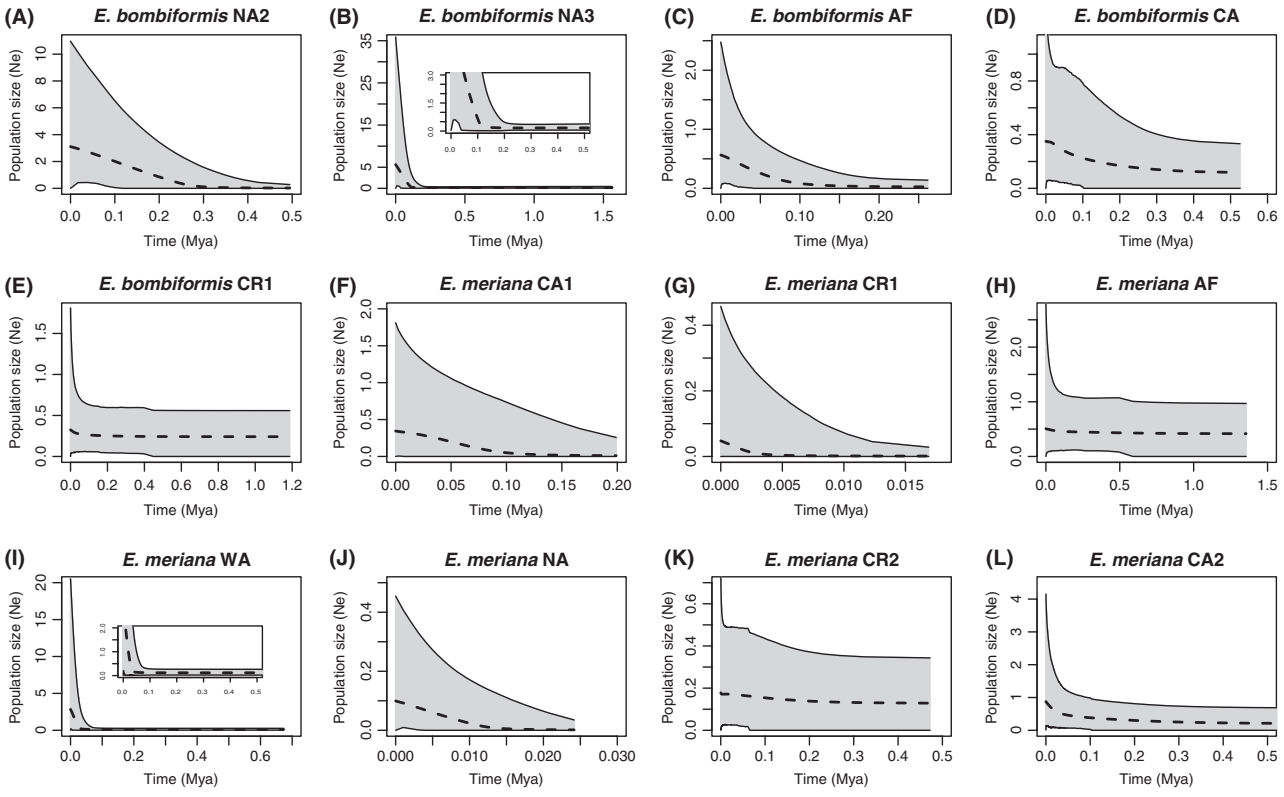

Fig. 4 Effective population size (*Ne*) through time (Mya) for monophyletic lineages in *E. bombiformis* (A–E) and *E. meriana* (F–L). Dotted lines indicate median *Ne*. Upper and lower lines indicate 95% highest posterior density intervals.

between geographical and genetic distance in all three focal species, a pattern absent in the nuDNA. Discordant cyto-nuclear patterns are common in recently diverged lineages (Gómez-Zurita & Vogler 2003; Bryja *et al.* 2010), and different processes may explain this pattern. First, discordance can result from long-term male-biased dis-

Table 5 Estimates of effective population migration rate for *E. bombiformis* from the IMa2 analyses. Values shown are peak probabilities with highest posterior densities at the lower and upper 95%

		$m_{0>1}$†			$m_{1>0}$‡		
Populations*	Genome	High Pt	HPD95Lo	HPD95Hi	High Pt	HPD95Lo	HPD95Hi
CR – CA	mt	0.0125	0.0075	97.49	0.509	0.193	89.16
	nu	0.0875	0.194	46 708	1.618	0.218	43 518
Am – AF	mt	0.0129	0.0061	48.55	0.015	0.0084	244.76
	nu	0.563	0.168	22 120	0.619	0.219	296.41

*Populations CR, CA, Am, AF correspond to clades 'CR2', CA, NA2 + NA3, 'AF' on Fig. 2A.
†Rate at which population 0 (CR, Am) receives genes from population 1 (CA, AF).
‡Rate at which population 1 (CA, AF) receives genes from population 0 (CR, Am).

persal resulting in panmixia in the nuDNA and genetic structure in the mtDNA. Alternatively, cyto-nuclear discordance can arise due to slower mutation rate and larger effective population size of the nuclear genome (Prugnolle & de Meeus 2002). Another possible explanation is that female traits, such as mtDNA, are adaptive resulting in mitochondrial haplotype frequencies selected by environmental conditions (Pavlova *et al.* 2013). In this study, we targeted two mitochondrial protein-coding markers and five nuclear markers with slow and fast mutation rates. Both types of nuclear markers, the slow-evolving protein-coding and the fast-evolving ASCN loci, lacked genetic structure, while the mtDNA loci consistently showed high population structure. Our sampling within CSAs included populations separated by more than 2500 km and with latitudinal variation of 23°. Therefore, adaptive selection seems an unlikely explanation for the cyto-nuclear discordance given the great environmental variation within sampled CSAs. Estimates of migration rates for *E. bombiformis* indicate lower gene flow between phylogeographical groups with the mitochondrial than the nuclear genome. These results do not completely eliminate the possibility that some of the observed patterns are due to incomplete lineage sorting, but they validate previous ecological and behavioural studies that suggested that gene flow is male-biased in orchid bees (López-Uribe *et al.* 2008).

Overall, palaeomodels and demographic inferences indicated that climatic instability, physiological tolerance and sex-biased dispersal shaped the current distribution of genetic diversity in our focal orchid bee species. Furthermore, our results corroborate that current wet forested areas in the Neotropics, particularly in the Amazon basin, were drier and supported smaller populations during the LGM (Solomon *et al.* 2008; Alfaro *et al.* 2012). This study provides clear evidence that climatic instability can be an important driver of intraspecific lineage diversification even in highly mobile organisms with large effective population sizes.

Phylogeographical history

Our mitochondrial results provide new insights into the origins and colonization histories of our focal orchid bee species (Ramírez *et al.* 2010). Because a single mated female bee can colonize a new area and establish a population (Zayed *et al.* 2007), the mtDNA is an informative marker to make inferences about colonization events even though this marker exclusively depicts the evolutionary history of maternal lineages. The mitochondrial topology for *E. bombiformis* indicates an origin in the Amazon Basin and later colonization west of the Andes. Our data indicate that the colonization of the Atlantic Forest from the Amazon basin occurred during a wet climatic period of the Pleistocene (~0.35 Mya; HPD = 0.28, 1.18) (Fig. 2A). The presence of two individuals from the northern Brazilian Amazon within the Atlantic Forest clade suggests a connection between these two regions through the Caatinga in northeastern Brazil (Wang *et al.* 2004; Batalha-Filho *et al.* 2012). Furthermore, a recent collection of *E. bombiformis* in mid-elevation forested islands in the middle of the Caatinga (Ubajara, Ceará, Brazil) (Nemésio & Ferrari 2012) supports the hypothesis of a forest connection between the Amazon and the Atlantic Forest through gene flow corridors in northern South America. We found a recent divergence between populations east and west of the Andes (~1.48 Mya; HPD = 0.88, 2.16) that post-dates the uplift of this Cordillera (Hoorn *et al.* 2010), indicating cross-Andean colonization rather than vicariance (Fig. 2A).

Eulaema meriana originated in Central America about ~3.7 Mya (HPD = 1.96, 5.8) and colonized the Chocó Region during the late Pliocene (~2.9 Mya; HPD = 1.61, 4.36) (Fig. 2B). Eastern South America was subsequently colonized after the formation of the Andes, and those populations diverged into two lineages [Atlantic Forest ('AF') and Amazon Basin (WA + CAm)], congruent with CSAs identified by the palaeomodels in that

region. During mid-Pleistocene, *E. meriana* colonized the Guiana's foothills in northern South American and recolonized the lowland forests west of the Andes where two different lineages diverged: one in the Chocó Region and one in Central America. Results from *E. meriana* corroborate the evidence of cross-Andean dispersal found in *E. bombiformis* and other orchid bee species with transcontinental distributions (Dick *et al.* 2004).

The lack of phylogenetic signal in the mtDNA of *E. cingulata* precludes inference of the phylogeographical history of this species (Fig. 2C). From the dating analysis, it is clear that *E. cingulata* has a more recent origin than *E. bombiformis* and *E. meriana.* Therefore, we interpret the absence of phylogeographical resolution for this species as one or a combination of the following: (i) incomplete lineage sorting due to the recent origin of the species, (ii) broader physiological tolerance to dry climatic conditions allowing gene flow through unsuitable habitats between CSAs and (iii) recent colonization by a maternal lineage from a single refugium. We caution that the estimated divergence times were based on mutation rates calibrated for mitochondrial genes in other insects. Therefore, these estimates should not be interpreted as absolute but relative times of divergence. Nonetheless, the low sequence divergence detected between lineages supports the recent dates estimated based on the mitochondrial mutation rates.

The phylogeographical groups we identified in this study are congruent with major biogeographical breaks (e.g. the Andes cordillera) that shaped the interspecific diversification of other widespread Neotropical taxa (Solomon *et al.* 2008; Martins *et al.* 2009). These studies report monophyletic lineages from Central America, the Amazon basin and the Atlantic Forests, but the presence/absence of lineages within these major phylogeographical areas is not concordant among all taxa examined to date. Phylogeographical studies in the Neotropics suggest that palaeoclimatic instability was an important driver of intraspecific diversification in highly mobile organisms as evidenced by the spatial and temporal congruence between genetic lineages and predicted CSAs (Solomon *et al.* 2008). However, general spatial patterns of vicariance and dispersal are difficult to predict due to taxon-specific physiological niche breadths that drive different responses to climatic changes.

Persistence and species vulnerability to future climate change

Our multilocus comparative study suggests that orchid bees have large effective population sizes and experience long-distance male-mediated gene flow that maintains high levels of genetic diversity and connectivity among populations. Our niche models show that the geographical distribution of these *Eulaema* species can be significantly reduced under dry climatic conditions, nonetheless, orchid bees seem resilient to the detrimental effects of climatic instability due to their long-distance dispersal capabilities and large effective population sizes. However, we observed different species-specific responses depending on physiological tolerance. Species with narrower niches are more susceptible to isolation as a result of climate change due to their inability to disperse through unsuitable habitat. In the face of continued climate change, those more susceptible species will be the first to suffer demographic consequences of geographical isolation due to reduction in suitable habitat. In addition, habitat degradation and severe reduction in wet forested areas due to agricultural intensification are likely to exacerbate the reduction in geographical distribution and increase in isolation. Our results shows that understanding species-specific responses to historical climate change, and combining those data with estimates of physiological tolerances, can predict which species will be most affected by future environmental changes.

Acknowledgements

We thank F. Campos, C. Keil and C. Castillo for their assistance with collecting permits; S. Ramírez, D. Roubik, S. Cardinal, M. Oliveira and M. Sánchez for providing museum specimens included in this study; L. Duque for his assistance in the field; S. Bogdanowicz, A. Green and C. Santiago for their assistance in the laboratory; and members of the Danforth lab, Zamudio lab, M. Oliveira and two anonymous reviewers for feedback on earlier versions of the manuscript. Funding was provided by grants from the Organization of Tropical Studies (OTS), Grace Griswold Endowment, Mario Einaudi Center, Lewis and Clark Exploration Fund, Explorer's Club Exploration Fund (to M.M.L-U), and by awards from the National Science Foundation (DEB-0814544 and DEB-0742998 to B.N.D.). Samples were collected and exported under the following permits: 614-2008 Costa Rica, 2009-027-QCAZ-E Ecuador, 304-2010 Peru and 48/2010 Brazil.

References

Alfaro JWL, Boubli JP, Olson LE *et al.* (2012) Explosive Pleistocene range expansion leads to widespread Amazonian sympatry between robust and gracile capuchin monkeys. *Journal of Biogeography*, **39**, 272–288.

Augusto SC, Garófalo CA (2011) Task allocation and interactions among females in *Euglossa carolina* nests (Hymenoptera, Apidae, Euglossini). *Apidologie*, **42**, 162–173.

Bartomeus I, Ascher JS, Wagner D *et al.* (2011) Climate-associated phenological advances in bee pollinators and bee-pollinated plants. *Proceedings National Academy Sciences USA*, **108**, 20645–20649.

Batalha-Filho H, Fjeldså J, Fabre P-H, Miyaki CY (2012) Connections between the Atlantic and the Amazonian Forest avifaunas represent distinct historical events. *Journal of Ornithology*, **154**, 41–50.

Bennett KD (1990) Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology*, **16**, 11–21.

Bonier F, Martin PR, Wingfield JC (2007) Urban birds have broader environmental tolerance. *Biology Letters*, **3**, 670–673.

Bryja J, Smith C, Konečný A, Reichard M (2010) Range-wide population genetic structure of the European bitterling (*Rhodeus amarus*) based on microsatellite and mitochondrial DNA analysis. *Molecular Ecology*, **19**, 4708–4722.

Carnaval A, Bates J (2007) Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in northeastern Brazil. *Evolution*, **61**, 2942–2957.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science*, **323**, 785–789.

Cheviron ZA, Hackett SJ, Capparella AJ (2005) Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. *Molecular Phylogenetics and Evolution*, **36**, 338–357.

Chown SL, Hoffmann AA, Kristensen TN, Angilletta MJ Jr, Stenseth NC, Pertoldi C (2010) Adapting to climate change: a perspective from evolutionary physiology. *Climate Research*, **43**, 3–15.

Crozier YC, Koulianos S, Crozier RH (1991) An improved test for Africanized honeybee mitochondrial DNA. *Experientia*, **47**, 968–969.

Danforth BN (1999) Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial COI sequence data. *Systematic Entomology*, **24**, 377–393.

Danforth BN, Ji S (1998) Elongation factor-1α occurs as two copies in bees: implications for phylogenetic analysis of EF-1α sequences in insects. *Molecular Biology and Evolution*, **15**, 225–235.

Danforth BN, Fang J, Sipes SD (2006) Analysis of family level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. *Molecular Phylogenetics and Evolution*, **39**, 358–372.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science*, **332**, 53–58.

Dick CW, Roubik DW, Gruber KF, Bermingham E (2004) Long-distance gene flow and cross-Andean dispersal of lowland rainforest bees (Apidae: Euglossini) revealed by comparative mitochondrial DNA phylogeography. *Molecular Ecology*, **13**, 3775–3785.

Ditchfield AD (2000) The comparative phylogeography of Neotropical mammals: patterns of intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. *Molecular Ecology*, **9**, 1307–1318.

Dressler RL (1982) Biology of the orchid bees (Euglossini). *Annual Review of Ecology and Systematics*, **13**, 373–394.

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

Eltz T, Sager A, Lunau K (2005) Juggling with volatiles: exposure of perfumes by displaying male orchid bees. *Journal of Comparative Physiology. A, Neuroetholy, Sensory, Neural and Behavioral Physiology*, **191**, 575–581.

ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, California.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Farrell BD (2001) Evolutionary assembly of the milkweed fauna: cytochrome oxidase I and the age of Tetraopes beetles. *Molecular Phylogenetics and Evolution*, **18**, 467–478.

Fjeldså J (1994) Geographical patterns for relict and young species of birds in Africa and South-America and implications for conservation priorities. *Biodiversity and Conservation*, **3**, 207–226.

Forister ML, McCall AC, Sanders NJ *et al.* (2010) Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. *Proceedings National Academy Sciences USA*, **107**, 2088–2092.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Gómez-Zurita J, Vogler AP (2003) Incongruent nuclear and mitochondrial phylogeographic patterns in the *Timarcha goettingensis* species complex (Coleoptera, Chrysomelidae). *Journal of Evolutionary Biology*, **16**, 833–843.

Heller NE, Zavaleta ES (2009) Biodiversity management in the face of climate change: a review of 22 years of recommendations. *Biological Conservation*, **142**, 14–32.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 183–195.

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

Hoorn C, Wesselingh FP, Ter Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Hugall A, Moritz C, Moussalli A, Stanisic J (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings National Academy Sciences USA*, **99**, 6112–6117.

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

Janzen DH (1971) Euglossine bees as long-distance pollinators of tropical plants. *Science*, **171**, 203–205.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics*, **6**, 13.

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Kieswetter CM, Schneider CJ (2013) Phylogeography in the northern Andes: complex history and cryptic diversity in a cloud forest frog, *Pristimantis w-nigrum* (Craugastoridae). *Molecular Phylogenetics and Evolution*, **69**, 417–429.

Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with

spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, **19**, 3727–3745.

Knowles LL, Carstens BC, Keat ML (2007) Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, **17**, 940–946.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

López-Uribe MM, Oi CA, Del Lama MA (2008) Nectar-foraging behavior of Euglossine bees (Hymenoptera: Apidae) in urban areas. *Apidologie*, **39**, 410–418.

López-Uribe MM, Green AN, Ramírez SR, Bogdanowicz SM, Danforth BN (2011) Isolation and cross-species characterization of polymorphic microsatellites for the orchid bee *Eulaema meriana* (Hymenoptera: Apidae: Euglossini). *Conservation Genetics Resources*, **3**, 21–23.

Maddison DR, Maddison WP (2005) *MacClade 4: Analysis of Phylogeny and Character Evolution*. Version 4.08a. http://macclade.org.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Martins FM, Templeton AR, Pavan ACO, Kohlbach BC, Morgante JS (2009) Phylogeography of the common vampire bat *(Desmodus rotundus)*: marked population structure, Neotropical Pleistocene vicariance and incongruence between nuclear and mtDNA markers. *BMC Evolutionary Biology*, **9**, 294.

Mulcahy DG, Morrill BH, Mendelson JR (2006) Historical biogeography of lowland species of toads (*Bufo*) across the Trans-Mexican Neovolcanic Belt and the Isthmus of Tehuantepec. *Journal of Biogeography*, **33**, 1889–1904.

Myers M, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nemésio A, Ferrari RR (2012) The species of *Eulaema (Eulaema)* Lepeletier, 1841 (Hymenoptera: Apidae: Euglossina) from eastern Brazil, with description of *Eulaema quadragintanovem* sp. n. from the state of Ceará. *Zootaxa*, **3478**, 123–132.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pavlova A, Amos JN, Joseph L *et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution*, **67**, 3412–3428.

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pimm SL, Kevan P (2000) Biodiversity: extinction by numbers. *Nature*, **403**, 843–845.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution*, **25**, 345–353.

Prado CPA, Haddad CFB, Zamudio KR (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology*, **21**, 921–941.

Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161–165.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramírez S, Dressler RL, Ospina M (2002) Abejas euglosinas (Hymenoptera: Apidae) de la región Neotropical: listado de especies con notas sobre su biología. *Biota Colombiana*, **3**, 7–118.

Ramírez SR, Roubik DW, Skov C, Pierce NE (2010) Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). *Biological Journal of the Linnean Society*, **100**, 552–572.

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

Roubik DW, Hanson PE (2004) *Orchid Bees: Biology and Field Guide*. INBio, San Jose, Costa Rica.

Sala OE, Chapin FS III, Armesto JJ *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Santos J, Coloma L, Summers K, Caldwell J, Ree R, Cannatella D (2009) Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. *PLoS Biology*, **7**, 448–461.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annual Review of Entomology*, **87**, 651–701.

Solomon SE, Bacci Jr M, Martins Jr J, Vinha GG, Mueller UG (2008) Paleodistributions and comparative molecular phylogeography of leafcutter ants *(Atta* spp.) provide new insight into the origins of Amazonian diversity. *PLoS ONE*, **3**, e2738.

Thomé MTC, Zamudio KR, Giovanelli JGR, Haddad CFB, Baldiserra Jr FA, Alexandrino JMB (2010) Phylogeography of endemic toads and post-Pleistocene persistence of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*, **55**, 1018–1031.

Trevan JW (1927) The error of determination of toxicity. *Proceedings of the Royal Society of London*, **101**, 483–514.

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

Wikelski M, Moxley J, Eaton-Mordas A *et al.* (2010) Large-range movements of Neotropical orchid bees observed via radio telemetry. *PLoS ONE*, **5**, e10738.

Zayed A, Constantin SA, Packer L (2007) Successful biological invasion despite a severe genetic load. *PLoS ONE*, **2**, e868.

M.M.L.-U. conceived the study, collected field and laboratory data, performed all data analyses and drafted the manuscript. K.Z. provided essential intellectual contributions and important feedback for data analyses and interpretation of results. C.F.C. provided specimens

from the Brazilian Atlantic Forest and assisted in manuscript preparation. B.N.D. funded the laboratory work and provided feedback on the manuscript.

Data accessibility

DNA sequences: GenBank Accession nos KF895558–KF970454. Final DNA sequence alignments, and MaxEnt input files were uploaded to Dryad: doi:10.5061/dryad.83 dm4.

Supporting information

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

Table S1 Locality data for all individuals included in this study.

Table S2 Summary table showing population assignment to CSA and sample sizes per population.

Table S3 Calculation of the relative niche breadth index (*RNB*) for *E. bombiformis*, *E. meriana* and *E. cingulata*.

Fig. S1 Habitat suitability binary models and sampling localities used for niche modeling analysis for *E. bombiformis* (A and D), *E. meriana* (B and E) and *E. cingulata* (C and F) based on current (A–C) and Last Glacial Maxima (D–F) environmental data. Red arrows indicate areas of under-prediction.

Fig. S2 Demonstration on how to calculate the relative niche breadth (*RNB*) index for *n* species (*i*) based on the response curves to *m* environmental variables (*j*).

Fig. S3 Spatial principal coordinate analysis (sPCA) of the first global score for the mitochondrial (A,C,E) and nuclear (B,D,F) data for *E. bombiformis* (A,B), *E. meriana* (C,D), and *E. cingulata* (E,F). Colored points denote individuals mapped onto the geographic space. Similar colors indicate similar genetic composition.

Fig. S4 Posterior probability densities for mitochondrial and nuclear mutation-scaled migration rates between populations from the east and west clades of *Eulaema bombiformis*. Parameters denote migration rate estimated as number of migrants per population. Populations CR, CA, Am, AF correspond to clades 'CR2', CA, NA2 + NA3, 'AF' on Fig. 2A.