

# Early and dynamic colonization of Central America drives speciation in Neotropical army ants

MAX E. WINSTON,\*† DANIEL J. C. KRONAUER‡ and CORRIE S. MOREAU†

\*Committee on Evolutionary Biology, University of Chicago, Chicago, IL 60637, USA, †Department of Science and Education, Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605, USA, ‡Laboratory of Social Evolution and Behavior, The Rockefeller University, New York, NY 10065, USA

## Abstract

The emergence of the Isthmus of Panama is one of the most important events in recent geological history, yet its timing and role in fundamental evolutionary processes remain controversial. While the formation of the isthmus was complete around 3 million years ago (Ma), recent studies have suggested prior intercontinental biotic exchange. In particular, the possibility of early intermittent land bridges facilitating colonization constitutes a potential mechanism for speciation and colonization before full closure of the isthmus. To test this hypothesis, we employed genomic methods to study the biogeography of the army ant genus *Eciton*, a group of keystone arthropod predators in Neotropical rainforests. Army ant colonies are unable to disperse across water and are therefore ideally suited to study the biogeographic impact of land bridge formation. Using a reduced representation genome sequencing approach, we show that all strictly Central American lineages of *Eciton* diverged from their respective South American sister lineage between 4 and 7 Ma, significantly prior to the complete closure of the isthmus. Furthermore, three of the lineage pairs form extensive and coincident secondary contact zones in Costa Rica and Nicaragua, with no evidence of gene flow. Such a discrete and repeated biogeographic pattern indicates at least two waves of army ant dispersal into Central America that were separated by significant genetic divergence times. Thus, by integrating phylogenomic, population genomic and geographic evidence, we show that early colonization of Central America across the emerging Isthmus of Panamá drove parallel speciation in *Eciton* army ants.

**Keywords:** biogeography, *Eciton*, Great American Biotic Interchange, Isthmus of Panama, macroevolution, reproductive isolation

Received 1 June 2016; revision received 22 August 2016; accepted 2 September 2016

## Introduction

Revealing historic change in geographic distributions is critical to understanding the origins and evolutionary dynamics of biodiversity (Jablonski *et al.* 2006). Range expansion can be as essential to the origins of regional biota as the speciation process itself, and therefore, proper models testing established biogeographic patterns should consider dispersal along with speciation and extinction (Rull 2011). Simultaneously, as landscape change has been viewed as the primary driver of

biological diversification to date, it is important to recognize that paleogeographic events likely affect evolutionary dynamics in several ways (Coyne & Orr 2004; Smith *et al.* 2014). For example, the uplift of the Isthmus of Panama (IP) and consequent closure of the Central American Seaway (CAS) had an unparalleled impact on biodiversity (Simpson 1980), oceanic currents, and global climate (Haug *et al.* 2005). Most notably, the IP provided a land bridge between two previously disconnected landmasses, prompting the Great American Biotic Interchange (GABI) (Stehli & Webb 1985).

As one of the most significant episodes of biological migration, the GABI is integral to decoding evolutionary dynamics of Neotropical and Nearctic biota (Bagley

Correspondence: Max E. Winston, Fax: 312 665 7754; E-mail: mewinsto@uchicago.edu

& Johnson 2014), yet many aspects of the timing and process of isthmus closure remain contentious (Stone 2013). For example, major differences in salinity and benthic carbon between the Pacific Ocean and the Caribbean Sea arising from 4.2 to 4.7 Ma (millions of years ago) suggest the end of deep water exchange and the CAS (Haug *et al.* 2001), and the measured shifts in oceanic currents and global climate suggest a full closure of the IP by 3.0–2.5 Ma (Bartoli *et al.* 2005). Conversely, recent geological discoveries of Panamanian fluvial deposits in South America argue for a closure of the CAS as early as *c.* 13–15 Ma (Montes *et al.* 2015). To add to this debate, the most thorough tectonic reconstructions of the region demonstrate the extant volcanic arc beginning to form at *c.* 12 Ma, with a series of islands clearly above sea level by 6 Ma (Coates *et al.* 1992). Evidence even shows that the corridors surrounding the Chorotega volcanic front in modern-day Nicaragua and Costa Rica emerged far before full closure, only to be submerged again for millions of years (Gutiérrez-García & Vázquez-Domínguez 2013). Although there is wide variation in the specific interpretations of this collective body of geological scholarship, the complexity of the emergence of the IP cannot be debated.

Interpretation of the biotic evidence and its relation to the complex geological substrate is perhaps more varied than the geological evidence of the emergence itself. Traditionally, the IP was considered a relatively simplistic land bridge that initiated the GABI after its emergence *c.* 3 Ma (Marshall *et al.* 1982). In this view, the isthmus provided the first ‘overland corridor’, designating taxa exchanged previous to this date as ‘heralds’ or ‘island hoppers’ (Stehli & Webb 1985). Although the original publications were faunally biased towards mammals, a thorough analysis of Caribbean and Pacific strata containing a multitude of near-shore marine fauna by Coates *et al.* (1992) strongly bolstered this argument with evidence suggesting a 3.5-Ma divergence date. Yet, the more recent treatment of the GABI in the literature has acknowledged spatial heterogeneity and temporal dynamics inherent to a complex rise of the IP. In fact, several clades—including predatory mammal families such as the Felidae, Mustelidae and Procyonidae—show early colonization of South America easily pre-dating the traditional timing of the rise of Panama (Woodburne 2010; Eizirik 2012). Furthermore, rather than treating the isthmus as a simple land bridge, a few recent studies have increasingly acknowledged the propensity of a complex geological substrate to generate diversity *in situ* through both dispersal and vicariant mechanisms via transient formation and disappearance of land bridges and islands during a transitional period of IP formation (Woodburne 2010; Bagley

& Johnson 2014). Lastly, and perhaps most strikingly, a reanalysis of hundreds of previous studies assuming a closure of 3 Ma showed that the most significant increase in intercontinental migration was around 8 Ma for several terrestrial clades (Bacon *et al.* 2015), suggesting that the traditional date of 3 Ma for the first land connection may be too late even for species that can disperse across aquatic barriers. Although these studies highlight variation in the inferred timing of species exchange across the IP, none of them explicitly consider dispersal strategies, which can have a profound effect on migration potential (Ree & Sanmartín 2009). Taxa that cannot disperse over water are particularly suitable for understanding the biogeographic history of land bridges as they produce fewer false positives for vicariant events (Cowie & Holland 2006), and are better indicators for evaluating landscape-driven speciation. We therefore chose a clade of Neotropical army ants to study the temporal and spatial dynamics of speciation across the IP, and to elucidate its complex geographic history.

Army ants of the genus *Eciton* are keystone arthropod predators exerting strong top-down trophic effects in terrestrial ecosystems (Kaspari & O’Donnell 2003; Kronauer 2009). Colonies of hundreds of thousands of individuals collectively hunt, kill and transport their prey (Kaspari & Vargo 1995). Due to this foraging strategy, new colonies can only form through fission, where a mature colony splits into two with each containing a single wingless queen (Kronauer 2009). Probably because of their massive colony sizes and predatory lifestyle, *Eciton* species have adapted a nomadic life history strategy to avoid local resource depletion (Willson *et al.* 2011). Although Neotropical army ant colonies cannot disperse over aquatic barriers due to wingless queens and workers that cannot swim or raft, gene flow is maintained by the dispersal of winged males across existing barriers (Berghoff *et al.* 2008). While male dispersal outperforms that of the queen and the rest of the colony, this male-biased dispersal also has its limits: genetic evidence for landscape effects on dispersal has been found in *Eciton burchellii* (Pérez-Espona *et al.* 2012; Soare *et al.* 2014), and simulations have suggested that extinction may occur far before the habitat is fully fragmented (Boswell *et al.* 1998).

Despite limitations in dispersal capability and range expansion, *Eciton* species assert their ecological dominance across broad geographic ranges in Neotropical rainforests, from Mexico to Argentina (Schneirla 1971; Watkins 1976). Previous work on the biogeographic origins of the multiple genera of army ants, including *Eciton*, suggests a Gondwanan origin followed by vicariant speciation resulting from the split of Gondwana into the South American and African continents (Brady

2003). Given this model and additional geographic evidence, *Eciton* species likely originated in the Amazon (Brady 2003; Scotese 2014) and colonized Central America during GABI following the closure of the IP (Stehli & Webb 1985; Hoorn *et al.* 2010). To tease apart the nature of this colonization process, we present two alternative models of colonization with explicit predictions, representing alternative interpretations of the geological and biological evidence on the emergence of the IP: The traditional Full Closure Colonization (FCC) model posits a simpler colonization of Central America beginning at full closure of the IP (c. 3 Ma), while the Early Dynamic Colonization (EDC) model posits a much earlier colonization (>3 Ma) across the spatiotemporally complex substrate of the emerging IP. Due to the weak dispersal capability of army ants (Schneirla 1971; Kronauer 2009), it is unlikely that colonization occurred before the completion of a single or a series of connected land bridges. This assumption is supported by the absence of Neotropical army ants from all noncontinental Caribbean islands, despite the ample ecological opportunity presented by these habitats (Gosti *et al.* 2007). Moreover, even with a rich and abundant fossil record in the Dominican amber deposits, no fossil *Eciton* have been found on the islands (Wilson 1985).

To test these alternative models of colonization across the IP, we generated a large, robust data set using a reduced representation genome sequencing approach, genotyping by sequencing (GBS). From these data, we infer the phylogeny of the genus *Eciton* and elucidate its biogeographic history. In addition to providing increased phylogenomic resolution compared with traditional gene-based methods (Eaton & Ree 2013), GBS facilitates population genomic analyses, which we leverage here to assess gene flow between distinct army ant lineages to test possible parallel, cryptic speciation. By collecting a large set of geographically distinct specimens for each species, we are able to test the predictions of two distinct colonization models—the FCC and the EDC—in generating observed Neotropical biodiversity.

## Materials and methods

### *Taxon sampling*

Taxa were selected for sequencing based on three main criteria: (i) broad taxonomic coverage within the New World army ants was required to assess congruence with previous phylogenetic work on Dorylinae (Moreau & Bell 2013; Brady *et al.* 2014); (ii) all nine *Eciton* species that are known from queens and workers were included to account for full taxonomic coverage; and (iii) extensive geographic coverage for all *Eciton* species—many of which (six of nine species) have cosmopolitan

Neotropical distributions extending from Mexico to Argentina—was required to study biogeographic patterns and to infer species boundaries. Comprehensive sampling of specimens across the geographic range was particularly important given the many taxonomically recognized and morphologically distinguishable subspecies (Watkins 1976) (Table S3, Supporting information). All voucher specimens have been deposited at the Field Museum of Natural History.

### *GBS library preparation and locus assembly*

DNA from all specimens was extracted following a standard protocol for ants (Moreau 2014). Library preparation for GBS followed a published protocol with a modified size-selection step (Elshire *et al.* 2011). The restriction enzyme *ApeK1* was chosen for genome digestion based on *in silico* digests of the eight published ant genomes available at the time of study design (Appendix S1: 1, Supporting information), and was later confirmed by an *in silico* digest and analysis (Appendix S1: 2, Supporting information) on a draft genome of *E. burchellii*. All libraries were size-selected for fragment sizes between 300 and 800 bp to maximize coverage and number of loci (Appendix S1: 1, Supporting information). Finally, three libraries with 50 bar-coded samples each were prepared for GBS by this protocol, totalling 150 specimens sent for single-end sequencing on three lanes of the Illumina HiSeq 2000. All sequences have been deposited in GenBank (SRA Accession: SRP072129). De novo locus assembly was accomplished using the data pipeline *pyRAD* (Eaton 2014). De novo locus assembly was a multistep computational process where sequencing reads were sorted and filtered for quality, and then clustered by sequence similarity within and between samples to form genomic loci with a minimum of 10 $\times$  coverage [ $\mu = 20.1x$ ] (Appendix S1: 3, Supporting information). The final set of genomic loci was then subset by (i) the minimum number of samples for which a locus genotype was available, (ii) the maximum number of individuals with a shared allele in the locus and (iii) the exclusion of specific taxa, depending on the particular analysis.

### *Phylogenomic inference*

Following quality filtering of the assembled GBS loci, all parsimony-informative loci were concatenated into a single data matrix with *pyRAD* for phylogenetic inference (Eaton 2014). This matrix included 135 *Eciton* samples and 11 samples from outgroup genera. Based on previous phylogenetic work in Dorylinae (Brady 2003; Brady *et al.* 2014), the three *Neivamyrmex* specimens in the data set were set as outgroups for rooting the tree

in both maximum likelihood (ML) and Bayesian inference (BI). ML inference was implemented in a Randomized Accelerated Maximum Likelihood with High Performance Computing (RAxML-HPC) pipeline with a GTR-GAMMA model and 100 rapid bootstrap trees (Appendix S1: 4, Supporting information) (Stamatakis 2006). BI was accomplished using an Exabayes MCMC approach with a GTR-GAMMA model of evolution and a Dirichlet prior (Appendix S1: 4, Supporting information) (Aberer *et al.* 2014).

#### Population genomic inference

Because large genomic data sets generated by GBS can lend strong phylogenetic support to populations that may not be truly distinct lineages due to gene flow, we chose to test our findings with population genomic inference. Specifically, gene flow between monophyletic groups as recovered by phylogenetic analyses was tested on a locus-by-locus basis. To manage the variable sample coverage between loci, we chose a modified Wright's  $F_{ST}$  estimator (Appendix S1: 5, Supporting information) weighted by sample size that could provide comparable  $F_{ST}$  estimates across our loci as defined in Chen *et al.* (2015).

As nucleotide diversity ( $\pi$ ) and sample coverage ( $n$ ) offer important evidence towards understanding potential gene flow, these statistics were also calculated and incorporated into the analyses, along with allele class, Hardy–Weinberg chi-squared statistic and minor allele frequency (Chen *et al.* 2015). As loci with singleton mutations are uninformative with respect to evaluating gene flow between two putative populations, they were removed from the analysis. Presence of large numbers of fixed loci, especially in loci with high nucleotide diversity and sample coverage, is a strong indicator that there is little to no gene flow between distinct lineages, despite geographic overlap in parapatric ranges. Loci with low nucleotide diversity may show reduced  $F_{ST}$  even in fully differentiated species if a novel allele has not spread through one of the species. Due to wide geographic sampling and population structure within the divergent parapatric lineages, distribution of  $F_{ST}$  in tested loci may deviate from normal two-population  $F_{ST}$  expectations. However, as the deviations would only reduce the sensitivity of our analysis by reducing  $F_{ST}$  values, no corrections are necessary to reduce type I error.

#### Tree dating

Conversion to an ultrametric tree and dating of the Bayesian inference (BI) phylogeny using node calibration was accomplished using the R package 'ape' (Paradis *et al.* 2004; Paradis 2013). We implemented the

*chronos* function (Paradis *et al.* 2004), which uses a penalized likelihood approach for estimating absolute rates of molecular evolution, known as nonparametric rate smoothing (NPRS) (Sanderson 2002). To choose the optimal value for the rate smoothing parameter ( $\lambda$ ), we used a twofold cross-validation step using the CV functionality in *chronos*, first cross-validating with a broad range of rate smoothing parameters ( $10^{-6} < \lambda < 10^6$ ), followed by finer resolution cross-validation ( $0.1 < \lambda < 2.0$ ). Once the optimal value was achieved, we leveraged the recent fossil-calibrated phylogeny for all Dorylinae using the 95% confidence interval for the marginal posterior probabilities of several generic nodes to date our tree (Appendix S1: 6, Supporting information) (Brady *et al.* 2014). Specifically, to test congruence between our inferred phylogeny and that of Brady *et al.* (2014), we used the marginal posterior probability distribution of the most recent common ancestor (MRCA) of *Eciton* and *Neivamyrmex* for calibration, and marginal posterior probability distributions of the MRCA of *Eciton* and *Nomamyrmex*, *Labidus* and *Cheliomyrmex* for validation (Brady *et al.* 2014). To ensure robustness in our results to differing calibrations, we followed the same procedure for the recent family-level dated phylogeny by Moreau & Bell (2013).

## Results

#### Locus assembly

Strict filtering of the demultiplexed Illumina reads for Phred quality, read length and adapter sequence resulted in 441.3 million utilizable reads for clustering across 147 samples, with a mean 3.0 million reads per individual specimen (Appendix S1: 3, Supporting information). Several steps of clustering and quality filtering, including a step for a minimum 10 $\times$  coverage for each locus at each individual, resulted in a mean of 58 095 loci per specimen. As is expected in reduced representation sequencing, the number of final loci was highly variable among specimens (Appendix S1: 3, Supporting information), which we modelled effectively with several linear models using processing statistics and phylogenetic distance (Appendix S1: 7, Supporting information).

#### Phylogenomic inference

Maximum likelihood (ML) and Bayesian inference (BI) resulted in congruent phylogenies with perfect statistical support for all but one species-level node within the genus *Eciton*, resolving these relationships within the genus with high confidence (Fig. 1; Appendix S1: 4, Supporting information). Both phylogenies included 135 *Eciton* samples and 11 samples from outgroup genera.

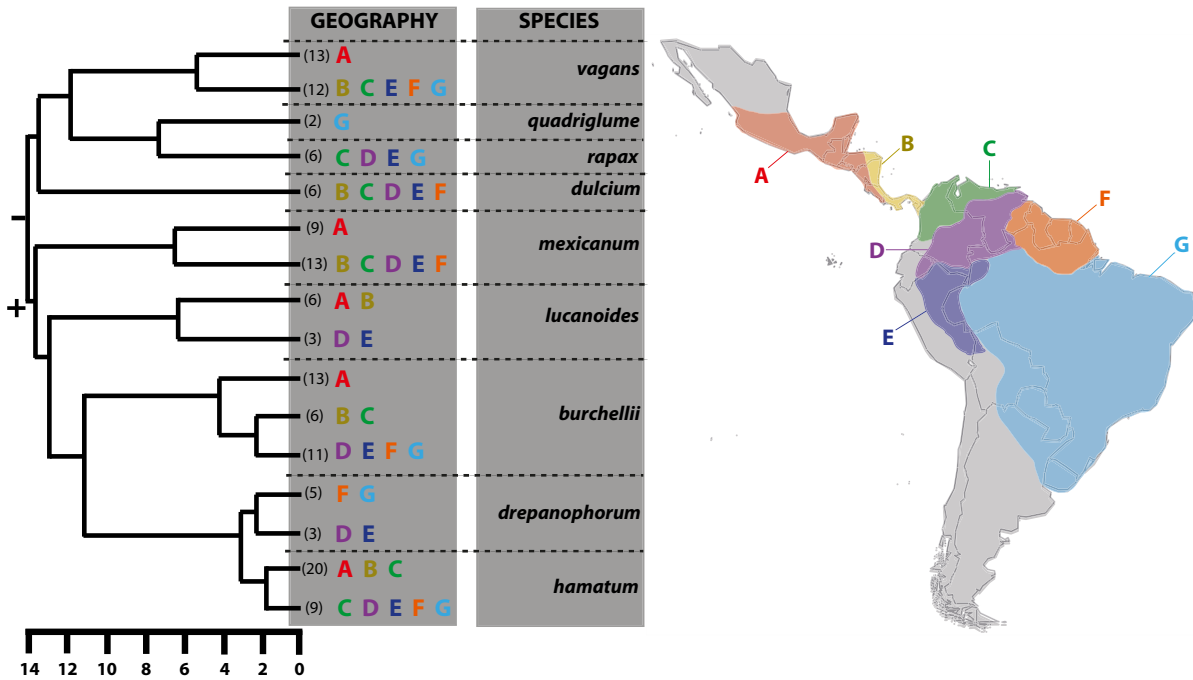


Fig. 1 The phylogeny of *Eciton* army ants and the biogeographic areas occupied by distinct lineages. Phylogeny inferred from 146 specimens and 419 804 loci with 6 700 494 distinct nucleotide sites. All nodes on the phylogeny had maximum support from ML bootstrap (BS) and Bayesian posterior probability (BPP) analyses, with one exception (+ = 95/100 BS, + = 0.98 BPP). Numbers in parentheses give the number of samples contained within each monophyletic group, and letters illustrate the biogeographic regions occupied by that group. Scale bar (bottom left) gives absolute timing of divergences in millions of years ago (Ma). Biogeographic areas are adapted from Morrone (2006) (51).

The generic-level topology of the tree (*Neivamyrmex* (*Cheliomyrmex* (*Labidus*, (*Nomamyrmex*, *Eciton*)))) confirms recent molecular phylogenetic work on Neotropical army ants (Brady 2003; Brady *et al.* 2014). Furthermore, all of the nine recognized *Eciton* species (Watkins 1976) were reciprocally monophyletic, validating current taxonomy (Fig. 1).

#### Biogeographic inference

Within each *Eciton* species, the phylogenetic signal was completely concordant with biogeography, in that each major monophyletic lineage occupied a distinct geographic area with minimal overlap with the areas of the other major monophyletic lineages from that species (Figs 1 and 3). Biogeographic areas were adapted from Morrone (2006), where any areas that lacked substantial occurrence data were merged into broader biogeographic areas of a more appropriate resolution. Mapping the occupied biogeographic areas for each sample onto the phylogeny, we found a remarkable amount of geographically coincidental, deep phylogenetic divergence across taxa, particularly around the Isthmus of Panama (Figs 1 and 3; Appendix S1: 4, 8, Supporting information). Three of the six species that have colonized Central America during the GABI had coincident

geographic boundaries between their respective oldest phylogenetically distinct lineages across the Chorotega volcanic front of Costa Rica into Nicaragua and Panama (Figs 1 and 3). In each case, the Central American lineage corresponded to a previously recognized subspecies (Watkins 1976). These three species with tightly coincident geographic boundaries—*E. burchellii*, *E. mexicanum* and *E. vagans*—are all known to have extensive secondary contact zones where their respective subspecies occur in parapatry (Fig. 3). The available biogeographic data suggest that this zone, which stretches throughout Costa Rica, Nicaragua and Panama, is 500 miles long and up to 100 miles wide (Gosti *et al.* 2007). For the fourth species (*E. lucanoides*) that has migrated into Central America, the two oldest lineages also meet at a geographic boundary along the Isthmus of Panama (Fig. 1; Appendix S1: 8, Supporting information). In the fifth species, *E. hamatum*, the two phylogenetically most distinct lineages meet along a geographic break near the junction of the isthmus and South America (Fig. 1; Appendix S1: 8, Supporting information). Lastly, *E. dulcium*, the sixth species that has colonized Central America, lacked distinct and deeply divergent intraspecific lineages (Fig. 1). However, it should be noted that our geographic sampling for this species was limited. In particular, we were not able to

include samples from north of the Chorotega volcanic front, although the species is known to extend into this region (Watkins 1976).

### Population genomic inference

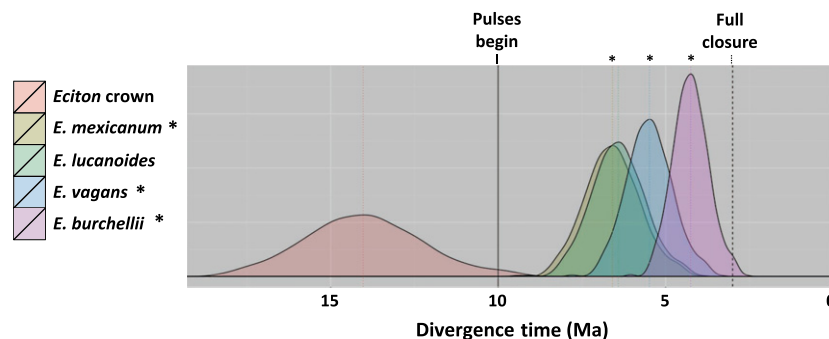
To further assess the hypothesis that the three discrete lineages of Central American *Eciton* army ants constitute distinct biological species, we gathered several locus-specific population genetic statistics (Appendix S1: 5, Supporting information). Nucleotide diversity ( $\pi$ ),  $F_{ST}$  (test and randomized pseudo-null) and sample coverage were calculated for a mean of 29 370 variable loci ( $\sigma = 11\,327$ ) from five pairs of distinct lineages associated with the IP, three of which were known to have extensive secondary contact zones in the Chorotega region (Appendix S1: 5, Supporting information). All five lineage pairs had several thousand loci ( $\mu = 11\,505$ ,  $\sigma = 5164$ ) with the maximum fixation index value ( $F_{ST} = 1$ ) throughout the genome (Appendix S1: 5, Supporting information). Limiting the analysis to variable loci with high nucleotide diversity ( $\pi > 0.8$ ) and high sample coverage ( $n > 9$ ) only increased this proportion, resulting in a mean of 55.0% ( $\mu = 4620$ ,  $\sigma = 2015$ ) that were fixed (Fig. 4; Appendix S1: 5, Supporting information). Finally, genotype assignments were randomized to create a pseudo-null distribution of  $F_{ST}$  values to compare to the results for each lineage pair. On average, less than three of the conservative loci were fixed for each lineage pair in the pseudo-null distribution. Therefore, the number of false positives expected by chance is orders of magnitude lower than the observed mean of 4620 fixed conservative loci across sister lineages, showing that the observed large number of fixed loci cannot be attributed to chance (Appendix S1: 5, Supporting information).

### Tree dating

We used a penalized likelihood approach with a twofold cross-validation step for divergence dating (Sanderson 2002; Paradis *et al.* 2004), calibrating with the marginal posterior probability for generic nodes from Brady *et al.* (2014). Specifically, using the most recent common ancestor (MRCA) of Neotropical army ants from Brady *et al.* (2014) as our fixed prior calibration, we estimated distributions for the MRCA of *Eciton* and the other Neotropical army ant genera for comparison with the remaining prior distributions from the same paper (Appendix S1: 6, Supporting information) (Brady *et al.* 2014). All of our estimated MRCA ages were within the original confidence intervals of Brady *et al.* (2014), suggesting general concordance across studies (Brady *et al.* 2014). Divergence date distributions were estimated for all five pairs of sister lineages across the IP (Fig. 2; Appendix S1: 6, Supporting information). Four of these lineage pairs had divergence dates significantly prior to the full formation of the isthmus and closure of the CAS (*E. burchellii*: 4.3 Ma [98.6% CI > 3 Ma]; *E. lucanoides*: 6.4 Ma [100.0% CI > 3 Ma]; *E. mexicanum*: 6.6 Ma [100.0% CI > 3 Ma]; *E. vagans*: 5.5 Ma [99.9% CI > 3 Ma]). The fifth parapatric lineage pair (*E. hamatum*) diverged 2.3 Ma, shortly after the full formation of the isthmus [98.6% PD < 3 Ma]. Calibration of our phylogeny with the date estimated by Moreau & Bell (2013) for the MRCA of Neotropical army ants yielded similar results, demonstrating that our findings are robust with respect to the precise date chosen for calibration (Appendix S1: 6, Supporting information) (Moreau & Bell 2013).

### Testing alternative colonization models

Testable predictions of colonization models rely heavily on the geological model of isthmus emergence



**Fig. 2** Divergence times for *Eciton* sister lineages across the Isthmus of Panama. Asterisks indicate lineages with geographically coincident species boundaries across Costa Rica, Nicaragua and Panama. Pulses of dispersal across the isthmus might have begun as early as 10 Ma (solid line) according to Bacon *et al.* (2015), while full closure of the Isthmus of Panama occurred around 3 Ma (black dashed line). All coloured dashed lines and areas represent the mean divergence times and divergence time density distributions estimated for the sister lineage of that colour, respectively.

considered. Here, we offer a null (FCC) model and an alternative (EDC) model, including their predictions based on the current state of the controversy in the literature and its implications for biological diversity. These models and their predictions are described below and summarized in Table 1.

First, we offer the Full Closure Colonization (FCC) model as our null model: traditionally, timing of the full formation of the Isthmus of Panama and closure of the CAS is dated at 3 Ma (Simpson 1980), and would represent the earliest date for a strictly over land dispersal into Central America. Under this model, we would expect dispersal of Neotropical army ant lineages into Central America to be within the last 3 Ma, with no particular expectation on where geographic range breaks may fall within Central America. Depending on the diversity within a given species group, we would also expect the earliest branching army ant lineages to be South American, and in situ diversification in Central America to be no older than 3 Ma (Table 1).

Second, we test our Early Dynamic Colonization (EDC) model, a model where the complex emergence of Panama acts as a mechanism for speciation on dispersing lineages: more recently, extensive biotic (Woodburne 2010; Bacon *et al.* 2015) and abiotic (Gutiérrez-García & Vázquez-Domínguez 2013; Montes *et al.* 2015) evidence has suggested that isthmus formation was far more sophisticated than originally proposed and that the emergence of suitable habitat and connective land mass may have predated the full closure of the CAS by several million years (Bagley & Johnson 2014). Under this model, dispersal into Central America presents both a novel ecological opportunity for colonization and an opportunity for clade diversification across a spatiotemporally complex landscape. Specifically, rather than a single migration across a suitable corridor into Central America, the complex landscape provided a series of patches of suitable habitat analogous to stepping stones with respect to the eventual colonization of Central America (Gutiérrez-García & Vázquez-Domínguez 2013). In this model, we would likely infer in situ diversification/speciation before full formation of the isthmus (*c.* 3 Ma) as a result of dispersing across a complex geological substrate. Lastly, secondary contact zones between parapatric sister-lineage pairs may be expected to coincide for species that diversified on the isthmus. Specifically, if these sister-lineage pairs shared the same geological mechanism of diversification—such as a large change in landscape providing a vicariance boundary—it is possible that the secondary contact zones would have remained coincident following the second wave of dispersal. On the other hand, given no known ecological differences between the respective members of each sister-lineage pair, there would be no reason to expect

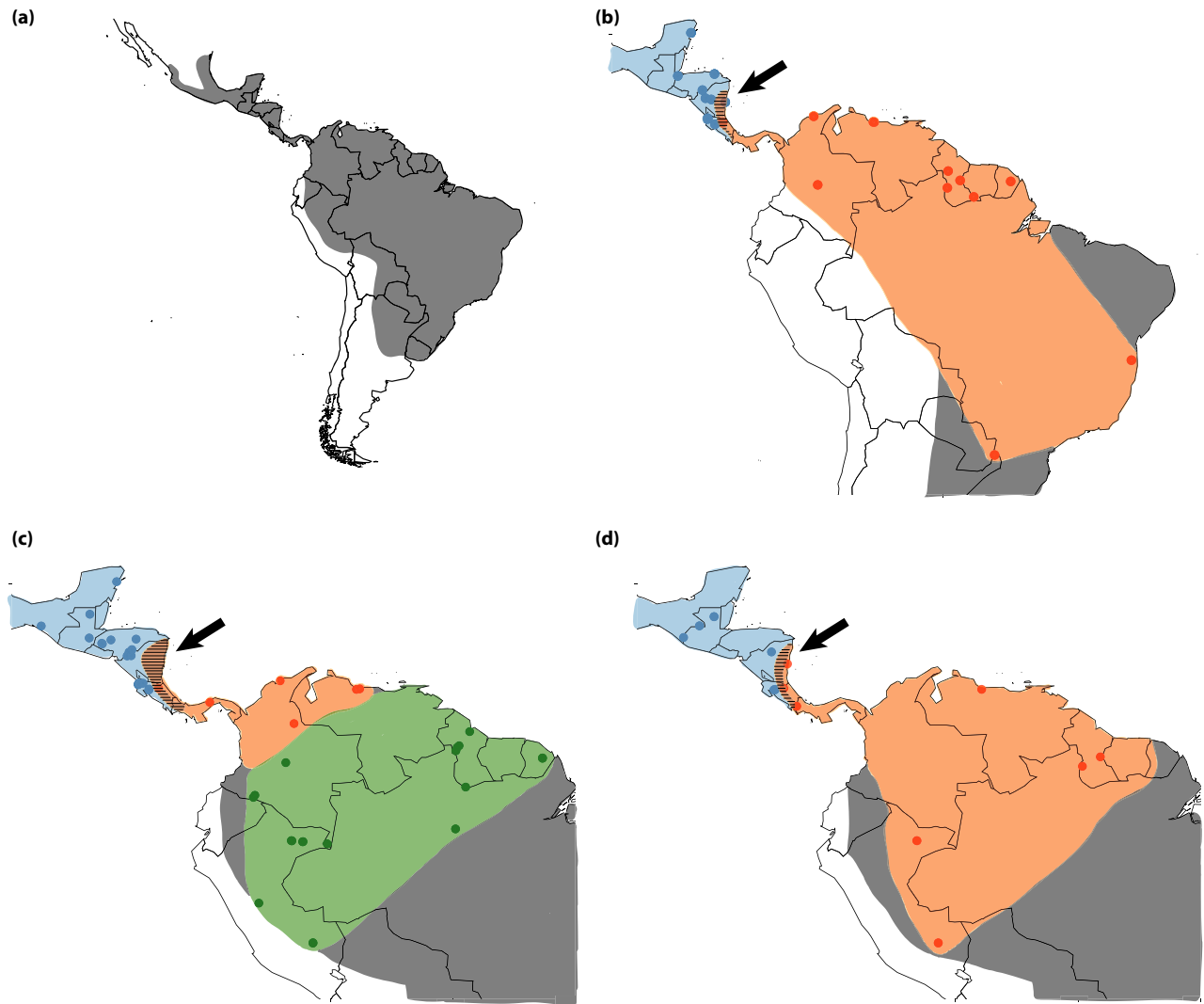
coincidence of secondary contact zones across species without a shared mechanism of diversification.

Overall, our results provide strong support for the predictions of the EDC model and reject the FCC model (Table 1). Specifically, we find that the majority of the diversification not associated with the early burst at the crown of *Eciton* occurs in the Central American region (Fig. 1), which is consistent with EDC, but not expected under FCC. In addition, for the three lineages that we can confirm as coincident and parapatric across the Chorotega volcanic range (*E. burchellii*, *E. mexicanum*, and *E. vagans*), their divergence times are all significantly older than 3 Ma [*E. burchellii*: 4.27; *E. vagans*: 5.50; *E. mexicanum*: 6.61] (Figs 2 and 3; Appendix S1: 6, Supporting information). The temporal and spatial congruence in the parapatric divergence of these lineages provides strong evidence supporting the EDC model. While *E. lucanoides* also has a deep phylogenetic break across the isthmus, this lineage is less abundant than the others, and thus, range parapatry cannot be ascertained at this point. *E. hamatum* also has a phylogenetic break near the Isthmus of Panama in Colombia. However, this divergence is younger than 3 Ma and requires increased geographic sampling to assess the extent of parapatry, as there are no subspecific designations in the literature setting the precedent for this break.

## Discussion

Understanding the factors that drive speciation is critical to revealing the origins of biodiversity and ecosystem assembly. Remarkably, we find strong evidence across multiple species for speciation associated with the complex emergence of Panama. In four of the six taxonomically recognized *Eciton* species with geographic ranges extending into Central America, our phylogenomic and population genomic data support the Early Dynamic Colonization model. In each case, the northernmost populations constitute a distinct lineage that is sister to the respective southern counterpart. In three of these cases—*E. burchellii*, *E. mexicanum* and *E. vagans*—we have identified coincident parapatric zones on the isthmus between the sister lineages that demonstrate evolutionary divergence and speciation far before the full closure of the isthmus (Figs 1–4; Appendix S1: 5 and 6, Supporting information).

Recent work on the uplift of Panama and the associated faunal exchange has continued to depart from the more simplistic view of a single event that provided a corridor for the GABI. Geologically, the tectonic reconstructions of volcanic island arcs (Coates *et al.* 2004), evidence for shoaling (Haug & Tiedemann 1998; Haug *et al.* 2001) and changes in sea level (Haug *et al.* 2005) provide an ample palette of mechanisms for speciation

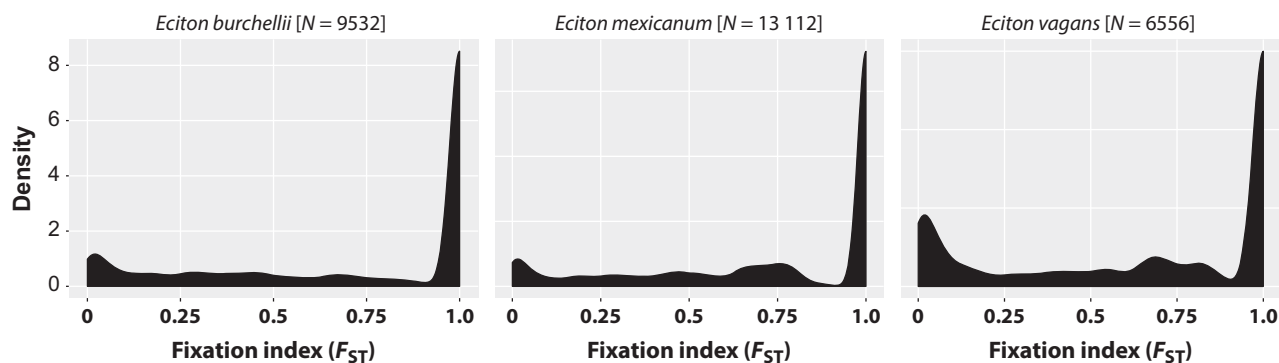


**Fig. 3** Geographic range for *Eciton* army ants with ranges and sampling for three species showing parallel speciation and coincidence of secondary contact zones. (a) The genus-level geographic range of *Eciton* is indicated by the grey shaded area, with most of the species in the genus occupying this cosmopolitan Neotropical distribution. Maps for each of the three species with distinct, overlapping lineages in Central America (b) *E. vagans*, (c) *E. burchellii* and (d) *E. mexicanum* with points representing the geographic coordinates of the specimens sequenced in this study. Colours indicate the assignment to a distinct phylogenetic lineage within each currently defined species, and black arrows indicate the secondary contact zones (hashed black lines) between the distinct lineages in northern Central America (blue) and southern Central America (orange) for each species. Although *E. burchellii* (c) also has a third distinct lineage (green), there are no data to suggest a secondary contact zone in South America. Estimated range areas (coloured by clade assignment) are based on our genomic data and geographic data associated with relevant, validated subspecies (Watkins 1976). Grey areas represent areas of the geographic range for that species that could not be assigned to a clade using genomic or geographic data. Note that many of the points represent multiple samples and some sites are obscured by the large scale of the map. For sample sizes, refer to Fig. 1.

via dispersal and vicariance. For groups such as Neotropical army ants that cannot disperse across water, it is likely that the continued exchange of shallow waters between the Pacific and Atlantic oceans sufficiently segmented the landscape during this time to create several 'holding pens' for evolutionary divergence (Woodburne 2010; Bacon *et al.* 2015). Under this scenario, colonization of the Central American landmass may have been an extremely slow process, with several

local extinctions of isolated island populations dispersing across the landscape, which may explain the deep divergence we find for all of the Central American lineages. Moreover, the high dietary demands and small population sizes of a top predator would only have exacerbated the propensity for local extinction, further impeding the colonization process (Reznick & Ricklefs 2009). Most importantly, geological reconstructions of the formation of the Central American volcanic





**Fig. 4** Distribution of  $F_{ST}$  values for conservative sets of loci for the three distinct lineage pairs with coincident secondary contact zones. Density distributions for *Ecton burchellii* (left), *E. mexicanum* (centre) and *E. vagans* (right). The number of loci tested for each species is indicated to the right of the species name above each plot. Note that the majority of loci for all three lineage pairs have the maximum  $F_{ST}$  values (1.0), as well as many other loci with elevated  $F_{ST}$  values ( $>0$ ).

**Table 1** Predictions made by the Full Closure Colonization (FCC) model and the Early Dynamic Colonization (EDC) model for various lines of evidence with model support from observed results. Evidence labels are as follows: Diversification is relative amount of speciation events associated with the IP compared to the remainder of the geographic range in the same time period. *Divergence times* refers to the divergence times of parapatric sister lineages with secondary contact zones in Central America, where we would expect younger ages for in situ diversification in FCC than EDC. *Parapatry coincidence* refers to the geographic coincidence of the secondary contact zones across parapatric sister lineage pairs, where we would potentially expect zone coincidence under EDC with a shared mechanism of diversification (<sup>†</sup>), and we see coincidence in our results for three lineages (<sup>‡</sup>)

Evidence	FCC prediction	EDC prediction	Results	Model support
Diversification	No effect	Increase	Increase	EDC
Divergence times	Less than 3 Ma	3–8 Ma	4–7 Ma	EDC
Parapatry Coincidence	Random	Yes <sup>†</sup>	Yes <sup>‡</sup>	EDC

archipelago reinforce our findings in both timing and specific path of dispersal, providing precise landscape features for colonization, as well as subsequent aquatic barriers for speciation (Appendix S1: 8, Supporting information) (Coates *et al.* 2004; Gutiérrez-García & Vázquez-Domínguez 2013).

When considering this evidence in the light of our alternative colonization models, we find clear support for the Early Dynamic Colonization (EDC) model over the Full Closure Colonization (FCC) model. Specifically, we find that the age of parapatric lineages found on the isthmus itself—all significantly older than 3.0 Ma—as strong evidence (Fig. 2), given that any assertion against

in situ diversification would require an unspecified mechanism of speciation that acted upon all of these lineages in parallel. Furthermore, any competing assertion would then require the migration of these boundaries hundreds of miles from South America into Central America in four independent lineage pairs, with three lineage pairs emerging as completely coincident along the Nicaraguan depression and Chorotega volcanic range (Fig. 3). As this scenario is highly unlikely, we thus reject the prospect of ex situ diversification. Additionally, it is clear that the majority of recent divergences (4 out of 7) are associated with the complex emergence of the IP, supporting the EDC model and indicating a clear increase in diversification around the IP in comparison with the much larger sampled geographic area in South America (Figs 1 and 3; Table 1; Appendix S1: 8, Supporting information). Although this support is not a formal statistical test, the concentration of divergences along the IP compared with the overall geographic area under study is striking.

As revealed by our population genomic data, the lack of gene flow in the extensive secondary contact zones of recently diverged lineages reinforces our argument for speciation (Fig. 4). More precisely, the fact that the majority of the loci in these lineage pairs demonstrate the maximum  $F_{ST}$  value ( $\mu = 11\,505$ ,  $\sigma = 5164$ ) despite our wide geographic sampling (Figs 1 and 3; Appendix S1: 8, Supporting information) would only be expected without gene flow (Appendix S1: 5, Supporting information). As the secondary contact zones between these lineages are large areas with ample opportunity for mating—on the order of thousands of square miles—these distributions of high  $F_{ST}$  values constitute one of the main lines of evidence for speciation. Combined with the comparison against our permutation-based pseudo-null that generated four orders of magnitude fewer loci with such extreme  $F_{ST}$  values

( $\mu = 2.6$ ,  $\sigma = 3.8$ ), we present strong evidence that these three co-occurring lineages are in fact distinct species (Appendix S1: 5, Supporting information).

Considering our EDC model for in situ diversification implicates large-scale abiotic processes associated with the rise of the isthmus rather than ecological specialization, competition among these species may be particularly strong in these regions (Price & Kirkpatrick 2009). Invading, ecologically similar lineages may thus have experienced competitive exclusion from closely related incumbent lineages, in addition to strong selective forces on premating isolation mechanisms to avoid costly hybridization (Coyne & Orr 2004). This perspective is particularly relevant in Neotropical army ants, where competitive exclusion seems to frequently occur between closely related sister lineages and niche partitioning appears common between more distantly related lineages: as demonstrated in the current study (Fig. 1), no ecologically similar sister lineages regularly co-occur in any part of their ranges. On the other hand, despite all species being top predators, more distantly related lineages partition niches in a number of ways—such as daily activity rhythms and food spectra—which may allow for co-occurrence without direct competition (Powell & Franks 2006).

The repeated pattern of deep divergences among Central American army ant lineages provides important new insights into the GABI and the closure of the IP. Aside from an independent source of validation for the mounting geological and biological evidence asserting an earlier and more complex emergence of the IP (Gutiérrez-García & Vázquez-Domínguez 2013; Stone 2013; Bagley & Johnson 2014; Bacon *et al.* 2015), our study shows this pattern in a clade without the ability to colonize land across aquatic barriers. As the early timing (4–7 Ma) of Central American colonization by Neotropical army ants was robust to multiple calibrations, this provides strong evidence for ephemeral land bridges prior to the full closure of the IP. We suggest that these early, intercontinental land connections played an important role in the dispersal of many taxa during the GABI (Gutiérrez-García & Vázquez-Domínguez 2013; Bacon *et al.* 2015), and likely functioned as a mechanism of speciation through alternating bouts of dispersal and vicariance. Considering that the majority of lineages in lower Central America exhibit genetic structure in this region (Bagley & Johnson 2014)—including other Neotropical army ant genera (Barth *et al.* 2015)—we predict that support for cryptic diversification and the EDC model will only increase as more research is conducted on weakly dispersing Central American species. Lastly, our research suggests that the spatial and temporal complexity of the emerging isthmus

needs to be seriously evaluated when invoking the commonly used biogeographic phenomenon of sweepstakes dispersal.

## Acknowledgements

This project was a highly collaborative effort, and would not have been possible without the generous researchers who sent specimens from across the Neotropics: Leanne Alonso, Fabricio Baccaro, Kaitlin Baudier, Christoph von Beeren, Séan Brady, Michael Branstetter, Alain DeJean, Fernando Fernández, Brian Fisher, Dan Janzen, Ana Jesovnik, John LaPolla, Jack Longino, Piotr Łukasik, Sean McCann, Sean O'Donnell, Scott Powell, Ben Rubin, Jeffrey Sosa-Calvo, Ted Schultz, Andrew Suarez, Phil Ward and Steve Yanoviak. Research was conducted in the Pritzker Laboratory for Molecular Systematics and Evolution at the Field Museum of Natural History, Chicago, Illinois, USA. For assistance in the laboratory, we thank Andrea Thompson and Brian Wray. This work was supported by the National Science Foundation (grant numbers NSF DEB1442316 to CSM and NSF DDIG1501672 to CSM and MEW). Lastly, we would like to thank an anonymous donor for a generous gift to support our work.

## References

- Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.
- Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6110–6115.
- Bagley JC, Johnson JB (2014) Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biological Reviews*, **89**, 767–790.
- Barth MB, Moritz RFA, Kraus FB (2015) Genetic differentiation at species level in the Neotropical army ant *Labidus* predator. *Insectes Sociaux*, **62**, 299–306.
- Bartoli G, Sarnthein M, Weinelt M, Erlenkeuser H, Garbe-Schönberg D, Lea DW (2005) Final closure of Panama and the onset of northern hemisphere glaciation. *Earth and Planetary Science Letters*, **237**, 33–44.
- Berghoff SM, Kronauer DJC, Edwards KJ, Franks NR (2008) Dispersal and population structure of a New World predator, the army ant *Eciton burchellii*. *Journal of Evolutionary Biology*, **21**, 1125–1132.
- Boswell GP, Britton NF, Franks NR (1998) Habitat fragmentation, percolation theory and the conservation of a keystone species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **265**, 1921–1927.
- Brady SG (2003) Evolution of the army ant syndrome: the origin and long-term evolutionary stasis of a complex of behavioral and reproductive adaptations. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 6575–6579.
- Brady SG, Fisher BL, Schultz TR, Ward PS (2014) The rise of army ants and their relatives: diversification of specialized

- predatory doryline ants. *BMC Evolutionary Biology*, **14**, 1–14.
- Chen G, Yuan A, Shriener D *et al.* (2015) An improved F(st) estimator. *PLoS One*, **10**, e0135368.
- Coates AG, Jackson JBC, Collins LS *et al.* (1992) Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin*, **104**, 814–828.
- Coates AG, Collins LS, Aubry MP, Berggren WA (2004) The Geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. *Geological Society of America Bulletin*, **116**, 1327.
- Cowie RH, Holland BS (2006) Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography*, **33**, 193–198.
- Coyne J, Orr HA (2004) *Speciation*. Sinauer, Sunderland, Massachusetts.
- Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
- Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: *Orobanchaceae*). *Systematic Biology*, **62**, 689–706.
- Eizirik E (2012) A molecular view on the evolutionary history and biogeography of Neotropical carnivores. *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals*, Vol. 3, pp. 123–143. University of Chicago Press, Chicago, Illinois.
- Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, 1937–1939.
- Gosti DA, Sanders NJ, Fitzpatrick MC *et al.* (2007) Global ant (Hymenoptera: Formicidae) biodiversity and biogeography – a new database and its possibilities. *Myrmecological News*, **77**–83.
- Gutiérrez-García TA, Vázquez-Domínguez E (2013) Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research*, **79**, 311–324.
- Haug GH, Tiedemann R (1998) Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature*, **393**, 1699–1701.
- Haug GH, Tiedemann R, Zahn R, Ravelo AC (2001) Role of Panama uplift on oceanic freshwater balance. *Geology*, **29**, 207–210.
- Haug GH, Ganopolski A, Signman DM *et al.* (2005) North Pacific seasonality and the glaciation of North America 2.7 million years ago. *Nature*, **433**, 821–825.
- Hoorn C, Wesselingh FP, Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.
- Jablonski D, Roy K, Valentine J (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–106.
- Kaspari M, O'Donnell S (2003) High rates of army ant raids in the Neotropics and implications for ant colony and community structure. *Evolutionary Ecology*, **5**, 933–939.
- Kaspari M, Vargo EL (1995) Colony size as a buffer against seasonality: Bergmann's rule in social insects. *The American Naturalist*, **145**, 610–632.
- Kronauer DJC (2009) Recent advances in army ant biology (Hymenoptera: Formicidae). *Myrmecological News*, 51–65.
- Marshall LG, Webb SD, Sepkoski JJ, Raup DM (1982) Mammalian evolution and the Great American Interchange. *Science*, **215**, 12.
- Montes C, Cardona A, Jaramillo C *et al.* (2015) Middle Miocene closure of the Central American Seaway. *Science*, **348**, 226–228.
- Moreau CS (2014) A practical guide to DNA extraction, PCR, and gene-based DNA sequencing in insects. *Halteres*, **5**, 32–42.
- Moreau CS, Bell CD (2013) Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution*, **67**, 2240–2257.
- Morrone JJ (2006) Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. *Annual Review of Entomology*, **51**, 467–494.
- Paradis E (2013) Molecular dating of phylogenies by likelihood methods: a comparison of models and a new information criterion. *Molecular Phylogenetics and Evolution*, **67**, 436–444.
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pérez-Espona S, McLeod JE, Franks NR (2012) Landscape genetics of a top neotropical predator. *Molecular Ecology*, **21**, 5969–5985.
- Powell S, Franks NR (2006) Ecology and the evolution of worker morphological diversity: a comparative analysis with *Eciton* army ants. *Functional Ecology*, **20**, 1105–1114.
- Price TD, Kirkpatrick M (2009) Evolutionarily stable range limits set by interspecific competition. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **276**, 1429–1434.
- Ree RH, Sanmartín I (2009) Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, **36**, 1211–1220.
- Reznick DN, Ricklefs RE (2009) Darwin's bridge between microevolution and macroevolution. *Nature*, **457**, 837–842.
- Rull V (2011) Neotropical biodiversity: timing and potential drivers. *Trends in Ecology & Evolution*, **26**, 508–513.
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, **19**, 101–109.
- Schneirla TC (1971) Chapter 4: Raiding. In: *Army Ants: A Study in Social Organization* (ed. Topoff HR), pp. 69–100. Freeman, San Francisco, California.
- Scotese CR (2014) *Atlas of Earth History, Vol. 1. Paleogeography*. PALEOMAP project, Arlington, Texas. 52 pp.
- Simpson GG (1980) *Splendid Isolation: The Curious History of South American Mammals*. Yale University Press, New Haven, Connecticut.
- Smith BT, McCormack JE, Cuervo AM *et al.* (2014) The drivers of tropical speciation. *Nature*, **515**, 406–409.
- Soare TW, Kumar A, Naish KA, O'Donnell S (2014) Genetic evidence for landscape effects on dispersal in the army ant *Eciton burchellii*. *Molecular Ecology*, **23**, 96–109.
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
- Stehli FG, Webb SD (1985) *The Great American Biotic Interchange*, pp. 357–386. Plenum Press, New York, NY.

- Stone R (2013) Battle for the Americas. *Science*, **341**, 230–233.
- Watkins JF (1976) *The Identification and Distribution of New World Army Ants*, pp. 1–109. Baylor University Press, Waco, Texas.
- Willson SK, Sharp R, Ramler IP, Sen A (2011) Spatial movement optimization in Amazonian *Eciton burchellii* army ants. *Insectes Sociaux*, **58**, 325–334.
- Wilson EO (1985) Ants of the Dominican amber (Hymenoptera: Formicidae). 2. The first fossil army ants. *Psyche*, **92**, 11–16.
- Woodburne MO (2010) The Great American biotic interchange: dispersals, tectonics, climate, sea level and holding pens. *Journal of Mammalian Evolution*, **17**, 245–264.

---

M.E.W. and C.S.M. designed the research; M.E.W. performed the research, analyzed the data, and wrote the paper; D.J.C.K., M.E.W. and C.S.M. provided samples and revised the manuscript.

---

### Data accessibility

*DNA Sequences*: GenBank Accession no SRP072129.  
*Concatenated Data Matrix*: Dryad Accession doi: 10.5061/dryad.3075h.  
*Reproducible Assembly Code*: Github repository (<https://github.com/mewinsto>).  
*Geographic and Sample Data*: Supporting Information.

### Supporting information

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

**Fig. S1** De novo locus assembly and quantitative modeling.

**Fig. S2** Full phylogeny inferred from both ML and Bayesian methods with bootstrap support and BPP.

**Fig. S3** Tree-dating procedure using nonparametric rate smoothing (NPRS) and absolute calibration and verification using credible intervals from Brady *et al.* (2014).

**Fig. S4** Plots of parameter estimates from converged, independent Bayesian inference MCMC chains.

**Fig. S5** Genetic relationships between the Central American and South American army ant lineage pairs.

**Fig. S6** Maps of clade assignments and geographic ranges for each Neotropical army ant.

**Appendix S1** Supplementary information text with eight subsections (SI. 1–8), supplementary figure legends (Fig. S1–S6), and supplementary table legends (Table S1–S3).

**Table S1** *In silico* digest results for ApeK1 on eight published ant genomes.

**Table S2** Partitioning of genetic diversity between parapatric pairs of sister lineages.

**Table S3** Sample information.