Insect Systematics and Diversity, 7(2), 1; 2023, 1–13
https://doi.org/10.1093/isd/ixad005
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular phylogenetics, phylogenomics, and phylogeography

Phylogenomics reveals within species diversification but incongruence with color phenotypes in widespread orchid bees (Hymenoptera: Apidae: Euglossini)

Stephania Sandoval-Arango[1,*], Michael G. Branstetter[2], Carolina F. Cardoso[3], Margarita M. López-Uribe[1,*]

[1]Department of Entomology, Center for Pollinator Research, Pennsylvania State University, University Park, PA 16802, USA,[2]U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS), Pollinating Insects Research Unit, Utah State University, Logan, UT 84322, USA,[3]Laboratório de Sistemática de Insetos, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil *Corresponding authors, mail: s.sandoval793@gmail.com (SS-A), mml64@psu.edu (MML-U)

Subject Editor: Jeffrey Lozier

Received on 13 September 2022; revised on 18 January 2023; accepted on 8 February 2023

Coloration is an important phenotypic trait for taxonomic studies and has been widely used for identifying insect species and populations. However, coloration can be a poor diagnostic character for insect species that exhibit high polymorphism in this trait, which can lead to over-splitting of taxonomic units. In orchid bees, color variation has been interpreted by different taxonomists as either polymorphism associated with Müllerian mimicry complexes or diagnostic traits for species identification. Despite this uncertainty, integrative approaches that incorporate multiple independent datasets to test the validity of hair coloration as a character that identifies independent evolutionary units have not been used. Here, we use phylogenomic data from Ultraconserved Elements (UCEs) to explore whether color phenotypes in the widespread orchid bee species complexes *Eulaema meriana* and *Eulaema bombiformis* (Hymenoptera: Apidae: Euglossini) correspond to independent lineages or polymorphic trait variation within species. We find that lineages within both species are structured according to geography and that color morphs are generally unassociated with evolutionarily independent groups except for populations located in the Atlantic Forest of Brazil. We conclude that there is compelling evidence that *E. atleticana* and *E. niveofasciata* are subspecies of *E. meriana* and *E. bombiformis*, respectively, and not different species as previously suggested. Therefore, we recognize *Eulaema meriana atleticana* **comb. n.** and *Eulaema bombiformis niveofasciata* **comb. n.** and discuss their morphological characteristics. We make recommendations on the use of color traits for orchid bee taxonomy and discuss the significance of subspecies as evolutionary units relevant for conservation efforts.

Key words: coloration, *Eulaema*, intraspecific variation, polymorphism, UCEs

Introduction

Coloration, either structural or due to pigments, is an important phenotypic trait for taxonomic studies and has been widely used for identifying insect species and populations (Shevtsova et al. 2011, Ghisbain et al. 2020). Color patterns have been especially important for the taxonomic identification of groups of insects that are morphologically challenging due to their small size (Perry and Heraty 2019) or that are structurally monotonous, exhibiting little to no morphological variation other than color (Ghisbain et al. 2020). Coloration, however, can also be a poor trait for delimiting and diagnosing species, in lineages that contain cryptic or polymorphic species. Using coloration in these cases can lead to both under- and over-estimation of biodiversity (Hines and Williams 2012, Ferrari and Melo 2014, Muñoz-Ramírez et al. 2016, Grando et al. 2018, Van Dam et al. 2022). Underestimation occurs in cryptic species because lack of color variation can lead to distinct species being

incorrectly lumped under one name (Chan and Grismer 2019). Conversely, over-estimation occurs in polymorphic species that exhibit significant variation in color because morphs are incorrectly split into taxonomic units (Grando et al. 2018, Lhomme et al. 2020, Van Dam et al. 2022).

In polymorphic species, color variation can have a simple or complex genetic basis whereby polymorphism is maintained via balancing selection (Plowright and Owen 1980, Baxter et al. 2010, Ando et al. 2018, Gautier et al. 2018). This mechanism is known to occur in insects that belong to Müllerian mimicry complexes (Williams 2007, Ferrari and Melo 2014, Grando et al. 2018). In these groups of species, intraspecific color variation can often be greater than interspecific variation, even for distantly related lineages [e.g., bumblebees (Hymenoptera: Apidae) (Hines and Williams 2012, Lecocq et al. 2015), and velvet ants (Hymenoptera: Mutillidae) (Wilson et al. 2015)]. In other cases, species with widespread, disjunct distributions can exhibit different phenotypes in allopatry and sympatry, complicating the interpretation of color variation for species identification [e.g., *Heliconius* butterflies (Lepidoptera: Nymphalidae) (Brown 1981, Kozak et al. 2015, Merrill et al. 2015)]. However, the alpha-taxonomy of some groups is still often defined by differences in setae or integument coloration because color variation can be involved in sexual selection and reproductive isolation (Gordon et al. 2018).

Due to the unreliability of using color variation alone to delimit and diagnose species, molecular data are being increasingly integrated into taxonomic studies to test morphology-based delimitations (Sites and Marshall 2004, Padial et al. 2010). Specifically, the use of phylogenetic inference from genome-wide markers in combination with morphological characters and life-history traits provides a powerful approach to resolve challenging taxonomic questions. A methodological framework based on independent datasets facilitates the interpretation of color variation in an evolutionary context to understand the processes behind the generation and maintenance of these color morphs. This approach has been powerful for resolving the taxonomy of enigmatic groups of ants (Longino and Branstetter 2020, Williams et al. 2022), parasitic wasps (Zhang et al. 2022), and bees (Lecocq et al. 2015, Freitas et al. 2018, Gueuning et al. 2020). Despite increasing evidence that color can be highly misleading in insect taxonomy, species continue to be diagnosed based mainly on coloration. One charismatic group where this is likely a problem is in the orchid bees (Nemésio 2009, Eltz et al. 2011, Ferrari and Melo 2014).

Orchid bees (Hymenoptera: Apidae: Euglossini) comprise a group of over 200 species endemic to the Neotropics with the highest number of species found in low-land wet forests (Roubik and Hanson 2004). The tribe comprises five genera: *Euglossa* (135 spp.), *Eufrisea* (70 spp.), *Eulaema* (29 spp.), *Exaerete* (9 spp.), and *Aglae* (1 spp.). Within the genus *E.* Lepeletier, several species have been described and formally named based largely on color variation (Moure 1967, 2000, Nemésio 2009). Some researchers, however, have disagreed with these delimitations due to lack of evidence and have lumped the various species/color morphs under species complexes, informal assemblages of phylogenetic related taxa that share morphological similarities, with unclear boundaries between them (Dressler 1979, Sousa-Paula et al. 2021). Within *Eulaema* species complexes, there is currently a lack of clarity on whether different color morphs represent different evolutionary units or if they represent distinct forms within single, polymorphic species.

In this study, we aim to integrate genomic and phenotypic information into the taxonomy of two species complexes of *Eulaema* orchid bees to shed light on their diversification and the evolution of coloration in each group. The first complex includes the species *Eulaema bombiformis* (Packard) and *Eulaema niveofasciata* (Friese) and is frequently named the '*E. bombiformis*' species complex. *Eulaema bombiformis* is distributed from Guatemala to the Amazon basin, and *E. niveofasciata* is restricted to the Atlantic Forest in Brazil. The second complex is named the '*E. meriana*' species complex and includes *Eulaema meriana* (Olivier), *Eulaema terminata* (Smith), *Eulaema flavescens* (Friese), *Eulaema quadrifasciata* (Friese), *Eulaema pallescens* Moure, and *Eulaema atleticana* Nemésio. *Eulaema meriana* is widely distributed across the Amazon basin, Choco region, and Central America. Although the names *E. quadrifasciata* and *E. pallescens* continue to be used in the orchid bee literature, they have been previously synonymized as morphs of *E. meriana* that occur in Costa Rica and Ecuador, respectively (Moure 1967, Nemésio 2009), and we treat them here as synonyms. *Eulaema terminata*, *E. flavescens,* and *E. atleticana* are restricted to Trinidad and Tobago, northern South America, and the Atlantic Forest, respectively. Species in both complexes exhibit aposematic colors (red, orange, yellow, and black) on the setae of the metasoma.

In a previous study, López-Uribe et al. (2014) investigated the phylogeographic history of these two species complexes and found multiple mitochondrial lineages that coincide with climatically stable areas during the Pleistocene. The congruence of these lineages with color phenotypes was not explored. Here, we combine molecular (UCEs) and morphological (coloration and body measurements) data to investigate whether the different color phenotypes within these two complexes are congruent with mitochondrial lineages previously identified and/or putative species boundaries (López-Uribe et al. 2014). UCEs are highly conserved genomic regions with adjacent variable regions (flanking DNA) that can be used to reconstruct the evolutionary history of taxa at multiple time scales (Zhang et al. 2019). For each individual in our UCE dataset, we recorded the coloration of the setae in the metasoma and measured a series of morphological characters that are important in the taxonomy of the focal species to analyze whether these non-color-related characters can separate lineages. As mentioned previously, the literature indicates that color polymorphism is common in insects and therefore our initial hypothesis is that phenotypic variation observed in *Eulaema* does not represent separately evolving lineages. Additionally, we compare phylogenetic reconstructions generated with mtDNA and UCEs to make inferences about the evolutionary history of *E. meriana* and *E. bombiformis* complexes and use these results for the interpretation of incipient species divergence in these groups of widespread Neotropical bees.

Materials and Methods

Study System and Sampling

The *E. meriana* and *E. bombiformis* species complexes exhibit aposematic colors (red/orange, yellow, and black) on their metasoma. Moure (1967, 2000) named several species and subspecies within these complexes based on coloration, but Dressler (1979) recognized only the nominal species, *E. bombiformis* and *E. meriana*, and described the existence of color morphs associated with mimicry rings that covary geographically (Fig. 1). While the *E. bombiformis* complex shows mainly two color patterns throughout its range, the *E. meriana* complex exhibits three frequent color forms in most of its distribution and only one in the Atlantic Forest (east coast of Brazil). In the Amazon basin, most individuals of *E. meriana* have a similar phenotype to *E. bombiformis* with hairs on the terminal segments displaying red

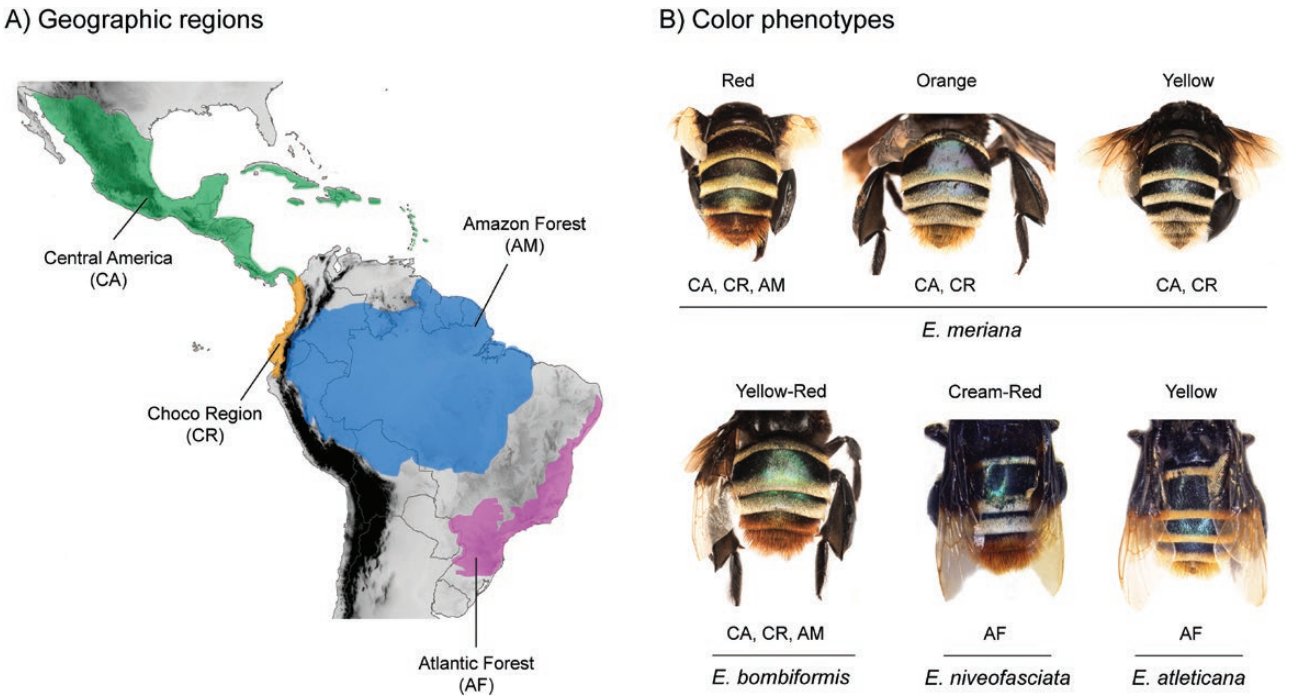

Fig. 1. Geographic distribution of the different color phenotypes in the *E. meriana* and *E. bombiformis* species complexes. (A) Colored areas in the map indicate approximate distribution for both species complexes as well as the different areas that correspond to lineages recovered in López-Uribe et al. (2014), including Central America (green), Choco region (orange), Amazon Forest (blue), and Brazilian Atlantic Forest (pink). (B) Photos of color phenotypes are shown for each species and the geographic region in which that phenotype is present. Photos of *E. meriana* and *E. bombiformis* by Nash Turley, photos of *E. niveofasciata* and *E. atleticana* by Marcelo de Oliveira Gonzaga.

coloration. West of the Andes (in the Choco Region and Central America), in addition to the red/orange morph, another *E. meriana* phenotype characterized by yellow bands on a black background is found in high frequencies. This phenotype is sometimes referred to as the '*E. flavescens* pattern' (subspecies *E. meriana flavescens*; Dressler 1979) but it is currently recognized as a separate species restricted to northern Venezuela (*E. flavescens*; Moure 2000) and has been named differently in other parts of the distribution such as Costa Rica (*E. quadrifasciata*) and Ecuador (*E. pallescens*). In this study, we included specimens from Costa Rica and Ecuador that have the 'flavescens' yellow pattern, but we were not able to include specimens from northern Venezuela. In the Atlantic Forest, on the east coast of Brazil, *E. atleticana* is colored similarly to *E. flavescens* but the yellow-colored bands on the metasoma are noticeably narrower (Nemésio 2009). Finally, *E. terminata*, although frequently included in this species complex, is restricted to Trinidad and Tobago and presents a very distinct morphology that clearly separates it from the other species (Moure 2000) and was therefore not included in this study.

To generate the UCE dataset, we selected 20 individuals from the *E. meriana* complex and nine from the *E. bombiformis* complex that represent the full range of the color variation in this group (Supplementary Table S1, Fig. 1), as well as lineages from across the distribution of the species that were recovered by López-Uribe et al. (2014) using mitochondrial data. These specimens were collected by López-Uribe et al. (2014) and permits for collecting are described in their methods. The mitochondrial lineages are distributed in the geographic regions highlighted on the map (Fig. 1). We selected a smaller number of individuals of *E. bombiformis* because the color patterns are more uniform across their distribution compared to *E. meriana*. In addition, following previous phylogenetic analyses of Euglossini (Cameron 2004, Ramírez et al. 2010) we included individuals of *Eulaema chocoana* Ospina and Sandino, and *Eulaema sororia* Dressler and Ospina-Torres, to improve resolution of the phylogenetic reconstruction within our ingroup and two individuals of *Eulaema cingulata* (Fabricius) were used as the outgroup.

Generation of the UCE Dataset

The UCE dataset was generated following the methodology described in Branstetter et al. (2021) combining targeted enrichment of UCE loci with multiplexed next-generation sequencing. The DNA extracts came from the study by López-Uribe et al. (2014), in which the front or hind legs of each individual were used for DNA extraction using the DNeasy blood and tissue extraction kit (QIAGEN) for fresh samples and phenol-chloroform for museum specimens (Danforth 1999). The concentration of the extracted DNA was measured using a Qubit 3.0 fluorometer (Thermo Fisher Scientific) and up to ~250 ng of genomic DNA was used as input into the UCE pipeline. Subsequently, library preparation, library pooling (10 samples per sequencing pool), UCE enrichment, qPCR quantification, final pooling (up to 110 samples per sequencing pool), and sequencing were performed. For UCE enrichment we used a recently published bait set that is specific to bees, ants, and other apoid wasps, but that targets loci inferred to be shared across the entire order Hymenoptera [hym-v2-bee-ant-specific; Grab et al. (2019)]. This bait set is a subset of the principal Hymenoptera bait set first reported in (Branstetter et al. 2017) and includes 9,068 unique baits targeting 2,545 UCE loci, plus an additional 264 baits targeting 12 exons from eight "legacy" markers (*ArgK, CAD, EF1a-F1, EF1a-F2, LwRh, NaK, POLD1, Top1*). Sequencing was performed on either an Illumina HiSeq 2500 (2 × 125 bp) at the University of Utah Genomics Core facility or on a HiSeq X (2 × 150 bp) at Novogene Inc, Sacramento, CA.

UCE Processing and Matrix Assembly

The data were cleaned, assembled, and aligned using PHYLUCE v1.6 [https://github.com/faircloth-lab/phyluce; (Faircloth 2016)], following a tutorial by Faircloth (2016) and the process outlined in Branstetter et al. (2021). Within PHYLUCE, we used Illumiprocessor v2.0 (Faircloth 2013) and Trimmomatic v0.4 (Bolger et al. 2014) for quality trimming raw reads, SPAdes v3.1 (Bankevich et al. 2012) for de novo assembly of reads into contigs, and LASTZ v1.0 (Harris 2007) for identifying UCE contigs from all contigs. For the LASTZ step, we specified the min-identity and min-coverage to 70 and 75, respectively, since these settings recover the highest number of UCE loci in bees when using the hym-v2-bee-ant UCE probes and bait file (Branstetter et al. 2021). To calculate assembly statistics, we used the PHYLUCE command *phyluce_assembly_get_fasta_lengths*.

Alignment of each UCE locus was performed using a stand-alone version of the program MAFFT v7.13 (Katoh and Standley 2013) and the L-INS-I algorithm. Subsequently, we used Gblocks v0.91b (Talavera and Castresana 2007) to trim flanking regions and poorly aligned internal regions, with the following parameters: b1:0.5, b2:0.5, b3:12, b4:7. Additionally, we used a PHYLUCE script to filter the initial set of alignments to test for differences in results due to different completeness matrices. We created three matrices requiring loci to have 75, 90, and 100% taxon completeness. For each set of filtered loci, we used the script *phyluce_align_get_align_summary_data* to generate descriptive statistics for the data, and *phyluce_align_format_nexus_files_for_raxml* to concatenate the loci into a supermatrix for analysis.

Phylogenetic Analyses

For the UCE dataset, we partitioned each of the concatenated matrices using the Sliding-Window Site Characteristics based on Entropy method [SWSC-EN; Tagliacollo and Lanfear (2018)]. This method partitions the UCE loci into three regions corresponding to core, right and left flank, and has been suggested as the best way to account for the UCE composition since core regions are conserved while flanking regions are variable. The resulting data subsets were merged using ModelFinder v2.0 when evolving under the same model within IQ-TREE v2.0 (Minh et al. 2020) using the rclusterf algorithm (checking only the top 10% of merging schemes), the corrected Akaike information criterion (AICc) and the GTR+G model. At this step we also calculated the models of sequence evolution for each matrix; in all cases, GTR+F+G4 was selected as the best model. We then performed maximum likelihood (ML) phylogenetic analyses for each matrix using IQ-TREE, performing 1000 ultrafast bootstrap replicates [UFB; Hoang et al. (2017)] and 1,000 replicates of the Shimodaira-Hasegawa (SH) approximate likelihood ratio test [aLRT; Guindon et al. (2010)].

Additionally, we inferred a Bayesian phylogeny of the mitochondrial dataset from López-Uribe et al. (2014) using only the taxa included in the UCE analysis to ease comparison between datasets. This includes the partial sequences of two mitochondrial regions, cytochrome oxidase subunit 1 (*CO1*, 825 bp) and cytochrome b (*Cytb*, 429 bp). The Bayesian phylogenetic reconstruction was conducted in BEAST2 v2.6.6 (Bouckaert et al. 2019) using the XSEDE computing cluster through the CIPRES Science Gateway v3.3 (Miller et al. 2010). We found the best-fit model of molecular evolution for each locus, including an analysis of different codon partitions based on the Bayesian Information Criterion implemented in jModelTest2 v2.1 (Darriba et al. 2012). We used Beauti to create the configuration file specifying HKY+I+G as the site model, a strict clock with a rate of 2.69% per million years (Papadopoulou et al. 2010) and a Markov Chain Monte Carlo (MCMC) run for 50 million generations saving trees every 1,000 generations. Parameter convergence was verified in Tracer v1.6 (Rambaut et al. 2018). We generated a maximum clade credibility tree (MCC) with mean heights and a 20% burn-in using TreeAnnotator v2.6.6 (Bouckaert et al. 2019). The final tree was visualized in FigTree v1.4.4 (https://github.com/rambaut/figtree/releases).

Finally, to explore the impact of individual UCE gene histories and the effects of incomplete lineage sorting on our dataset, we used *Beast (Heled and Drummond 2010) in BEAST2 v2.6.6 to estimate a species tree under the Multi-Species Coalescent model (MSC). Due to computational constraints associated with the parameter space in species tree analyses, we selected the 50 most informative UCE loci from our 100% completeness matrix using the PHYLUCE script *phyluce_align_get_informative_sites*. The individual nexus alignments of each UCE locus were then loaded into BEAUti v2.6.6 (Bouckaert et al. 2019) as individual unlinked genes. The taxon sets were set up with each individual as its own putative species so that the analysis was not biased by predefined species names. GTR+G4 was used as the site model using empirical frequencies. For the clock model, we selected a strick clock and asked the program to estimate the clock rate. We specified a multispecies coalescent model with a population mean = 1, pop function = linear with constant root, and all genes were specified as autosomal nuclear. We used a Yule model with default parameters as priors. We ran eight replicates of 200 million MCMC sampling every 10,000 generations using the CIPRES Science Gateway v3.3. Parameter convergence of each independent run and joint runs were verified in Tracer v1.6. We combined the tree files using LogCombiner v2.6.6 (Bouckaert et al. 2019) and generated a MCC tree with mean heights and a 20% burn-in using TreeAnnotator v2.6.6 (Bouckaert et al. 2019). The final tree was visualized using FigTree v1.4.4.

Estimation of Divergence Times

We inferred a time calibrated tree using BEAST2 v2.6.6. Because dating phylogenies is computationally intensive, we simplified the analysis by using a subset of the total available UCE loci and a constraint tree. We selected 500 random UCE loci from the 100% completeness matrix using the PHYLUCE script *randomly_sample_and_concatenate*. We then concatenated the 500 loci into a single supermatrix without partitioning. For the constraint tree, we used the topology obtained in our ML analysis of the 100% completeness matrix alignment. Before providing the tree to BEAUti, we converted it to an ultrametric tree using the *chronopl* function in the R package "ape" v5.6-2 (Paradis et al. 2004), which reduced the chances that our analysis would fail during the initialization phase of each run. To make our tree ultrametric and for the subsequent dating analyses, we used a secondary calibration from a re-analysis of Ramírez et al. (2010) dataset based on Sandoval-Arango (2018) (Supplementary Fig. S1). This was necessary because there are no fossils of *Eulaema*, but fossils of other orchid bees have been described. We placed our calibration point at the node that connects the ingroup and *E. cingulata*, using a normal distribution with a mean (*M*) set to 12 Mya and a confidence interval (*S*) of 0.5. We used a GTR+G4 substitution model with empirical frequencies, a relaxed log normal clock model and the birth-death model for the tree prior. Most parameters in the prior were set as default except for the ucldMean parameter which we specified as exponential. All the tree topology search parameters were set to zero, and we set starting values for the ucldMean and ucldStdev to 0.001 and 0.01, respectively. For each run, we set the MCMC length to 200 million generations, sampling

every 5,000 generations. We ran a total of six independent replicates using CIPRES Science Gateway v3.3. Parameter convergence verification as well as the generation of the MCC tree was performed as in our *Beast analysis.

Morphological Analyses and Visualization

We used two different approaches to analyze the morphological variation of characters that are associated with species identification in both complexes. First, we collected the morphological measurements from additional individuals in the *E. bombiformis* (n = 70) and *E. meriana* complexes (n = 197): body length (BL), head width (HW), intertegular distance (ID), width of colored bands on tergum II (WTII) and width of colored bands on tergum III (WTIII) (Supplementary Tables S2 and S3). We chose these morphological characters because they are used to differentiate species within the two complexes (Nemésio 2009). We visualized the morphological divergence between individuals grouped based on their geographic distribution using a multivariate Principal Component Analysis (PCA). Significant differences between groups were tested using one-way ANOVA and Tukey's honest tests. All statistical analyses were conducted in the R package "agricolae" v1.3-5 (De Mendiburu and Yaseen 2020).

Second, we mapped the measured characters on both phylogenies obtained from mtDNA and UCEs to determine whether there is a correlation between the morphological variation and the phylogenetic clusters. We also included the coloration of the setae on tergum V (T-V) in the analyses to observe whether different morphs (yellow, orange, and red) group together or separately in the phylogeny and if we can infer when and how often each color pattern has evolved (color patterns in Supplementary Table S1). We pruned the tree to include only *E. meriana* and *E. atleticana* because *E. bombiformis* and *E. niveofasciata* present only one color phenotype each in our dataset. We used the R package "phytools" v1.0-3 (Revell 2012) to plot the measured characters and color phenotypes in a comparative way using the Phyloheatmap function.

Results

Phylogenetic Relationships Within the *E. bombiformis* and *E. meriana* Complexes

Our mtDNA dataset included a total of 1,272 bp of which 159 bp were variable. In the UCE dataset, we recovered an average of 41,950,575 bp per individual after the SPAdes assembly. Following the assembly and extraction of UCE contigs, we aligned the data using MAFFT and trimmed the sequences using Gblocks, resulting in 2415 UCE loci with a mean alignment length of 845.93 bp per loci, a total of 49,602 informative sites, and an average of 20.54 informative sites per loci. The UCE data matrices included 2205, 2139, and 2022 loci for the 75, 90, and 100% taxon completeness, respectively. Detailed sequencing and assembly statistics can be found in Supplementary Table S4. We found no differences between the IQ-TREE results of the three completeness matrices (Supplementary Figs. S2 and S3), therefore subsequent analyses were performed on the 100% matrix and datasets.

For both molecular datasets, we recovered several monophyletic lineages that are mostly congruent with geography (Fig. 2, Supplementary Figs. S4 and S5), and with previous results based on mitochondrial data only (López-Uribe et al. 2014). These geographic clades have high support values and only some internal clades show low support (indicated by asterisks). However, we found incongruences between the two datasets, particularly for *E. meriana*. The mitochondrial Bayesian phylogeny recovered six main clades for the *E. meriana* complex, while the UCE ML phylogeny from IQ-TREE recovered three main clades for the *E. meriana* complex (Fig. 2). The main difference in the topology is explained by a split of the individuals corresponding to the Choco and Central American regions in the mitochondrial dataset (Fig. 2A). In both cases, *E. meriana* bees from the Amazon and *E. atleticana* bees from the Atlantic Forest formed one monophyletic group sister to either a clade that contains all the Central American individuals (Fig. 2B) or to both Central American and Choco individuals (Fig. 2A). For the *E. bombiformis* complex results were very similar between the two datasets. Individuals of *E. niveofasciata* formed a monophyletic group that was either sister to all the *E. bombiformis* individuals (Fig. 2A) or that was clustered in a clade that contains both *E. bombiformis* and *E. niveofasciata* bees (Fig. 2B).

Our species tree analysis included 50 UCE loci and recovered a topology similar to that of our IQ-TREE analyses with some incongruences in the internal nodes of most clades (Fig. 3). Within the *E. meriana* complex, there is support for an independent lineage that contains all the individuals from the Choco Region. Although without support, most individuals in the other clades grouped according to their geographic regions, except for one individual of *E. meriana* from Central America (ML172). The *E. atleticana* individuals were grouped with individuals from the Amazon Forest similar to our IQ-TREE analysis. In the *E. bombiformis* complex, the only clade that was supported by most trees was a lineage containing all the *E. bombiformis* and *E. niveofasciata* individuals. Clades within that lineage had low support, signaling incongruence between the species trees from independent runs and the gene trees of individual UCE loci.

The results of our divergence dating (Fig. 4) indicate that most diversification within the *E. meriana* complex occurred during the late Pliocene and early Pleistocene (Mean age: 3.88 Mya, 5–2.5 Mya 95% HDP). Additionally, the MCRA of the chocoan lineage of *E. sororia* and *E. chocoana* separated from the *E. meriana* complex approximately in the late Miocene [Mean age: 7.09 Mya, 9.5–4.5 Mya 95% highest posterior density (HDP)]. Finally, the *E. bombiformis* complex is likely to be an older lineage (7.5–3 Mya 95% HDP) compared to *E. meriana*, which radiated at the end of the Miocene and early Pliocene.

Coloration and Morphological Measurements

For the *E. meriana* complex, mapping color phenotypes onto the phylogeny shows that coloration does not separate well-defined groups for neither the mtDNA (Fig. 5A) nor the UCE phylogeny (Fig. 5B). Instead, individuals with yellow, orange, and red colors can be observed across clades. However, variation in the color of the last tergites of the metasoma was higher in the lineages from Central America and Choco, while specimens from the Amazon were completely red, and the Atlantic Forest specimens were completely yellow as described previously in the literature (Dressler 1979). For the continuous characters, we did not find clear groups based on morphology that coincide with the phylogenies. *Eulaema atleticana* individuals have narrower bands on tergum III (WTIII) when compared to *E. meriana* from other regions, and their body size measurements of the head width (HW) and intertegular distance (ID) are larger than those of individuals from other regions.

To investigate if individuals from the different geographic regions were morphologically different based on quantitative measurements, we performed an analysis of variance of the scaled PCA that showed significant morphological differences between the groups outlined

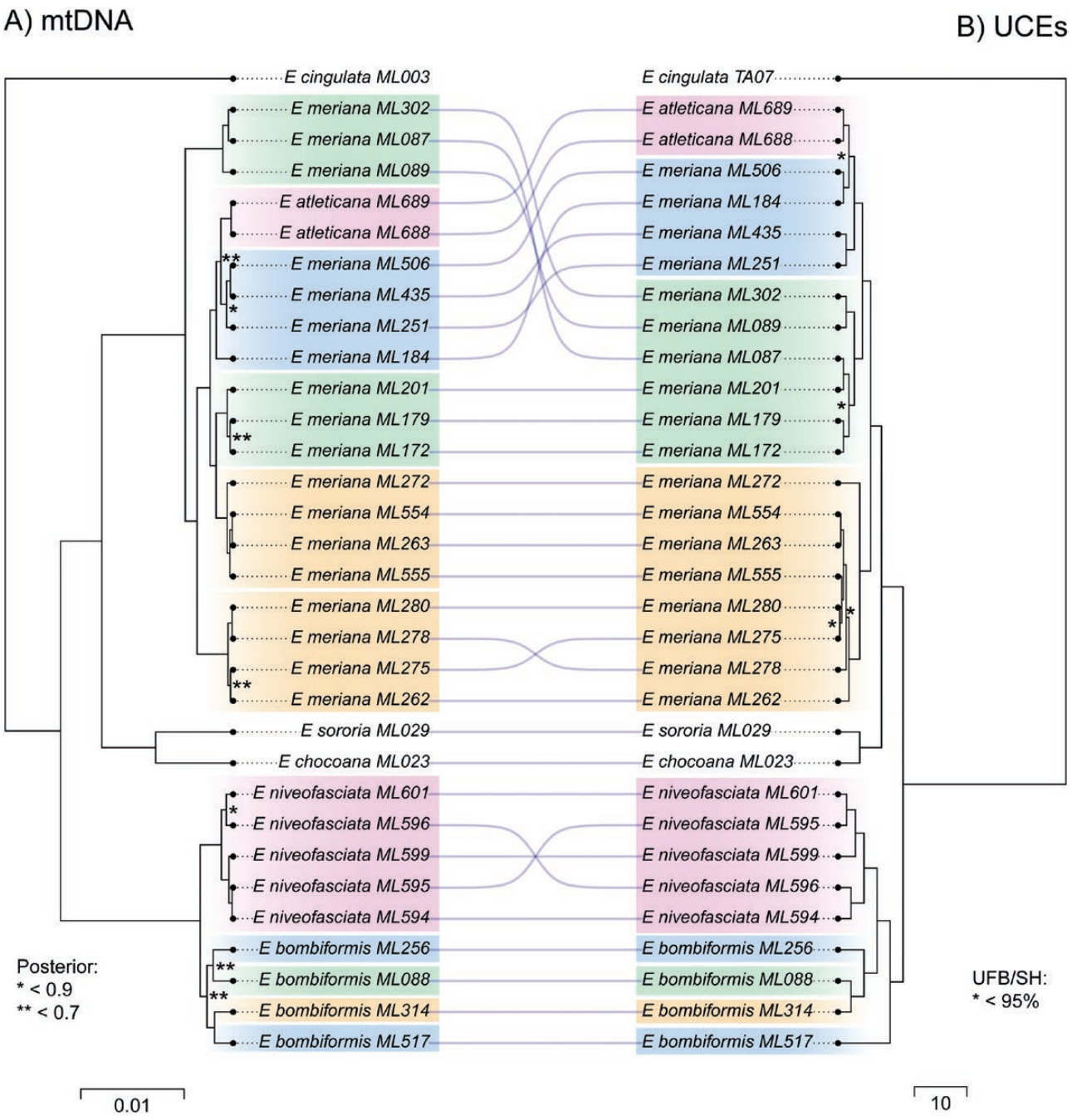

Fig. 2. Phylogenetic relationships of the *E. meriana* and *E. bombiformis* species complexes based on (A) mitochondrial data (mtDNA; *CO1* and *Cytb*) and (B) ultraconserved elements (UCE; 2022 loci). The mtDNA phylogeny was estimated using Bayesian inference in BEAST2, posterior probabilities on nodes were all above 0.9 except for nodes with asterisks (*). The UCE phylogeny was estimated with maximum likelihood using IQ-TREE and a concatenated 100% completeness matrix. Support values on nodes indicate ultrafast bootstrap (UFB) and SH-like (SH) approximate likelihood ratio test scores (SH-aLRT). All support values were above 95/95 except for nodes indicated with asterisks (* or **). One of the *E. cingulata* individuals (TA12) was pruned to improve the cophylogenetic visualization. The colors of the clades correspond to geographic regions outlined in Fig. 1. Green: Central America, Orange: Choco region, Blue: Amazon Forest, and Pink: Atlantic Forest.

by different regions (*E. bombiformis*: $F_{3,66} = 21.96$, *P*-value < 0.001; *E. meriana*: $F_{3,193} = 5.79$, *P*-value< 0.001; Fig. 6). These quantitative morphological measurements indicated three clearly differentiated morphological groups: (i) individuals of *E. bombiformis* from the Choco region and Central America (Fig. 6A) and (ii) of *E. atleticana* from the Atlantic Forest (Fig. 6B). The strongest morphological differentiation was supported by the width of the color bands on TII and TIII. For *E. bombiformis* Choco region, individuals showed broader color bands on these tergites. *E. atleticana* individuals had narrower color bands on TII and TIII when compared to *E. meriana* individuals.

Discussion

Our results indicate that there is incongruence between color variation, morphological traits, and the phylogenetic relationships

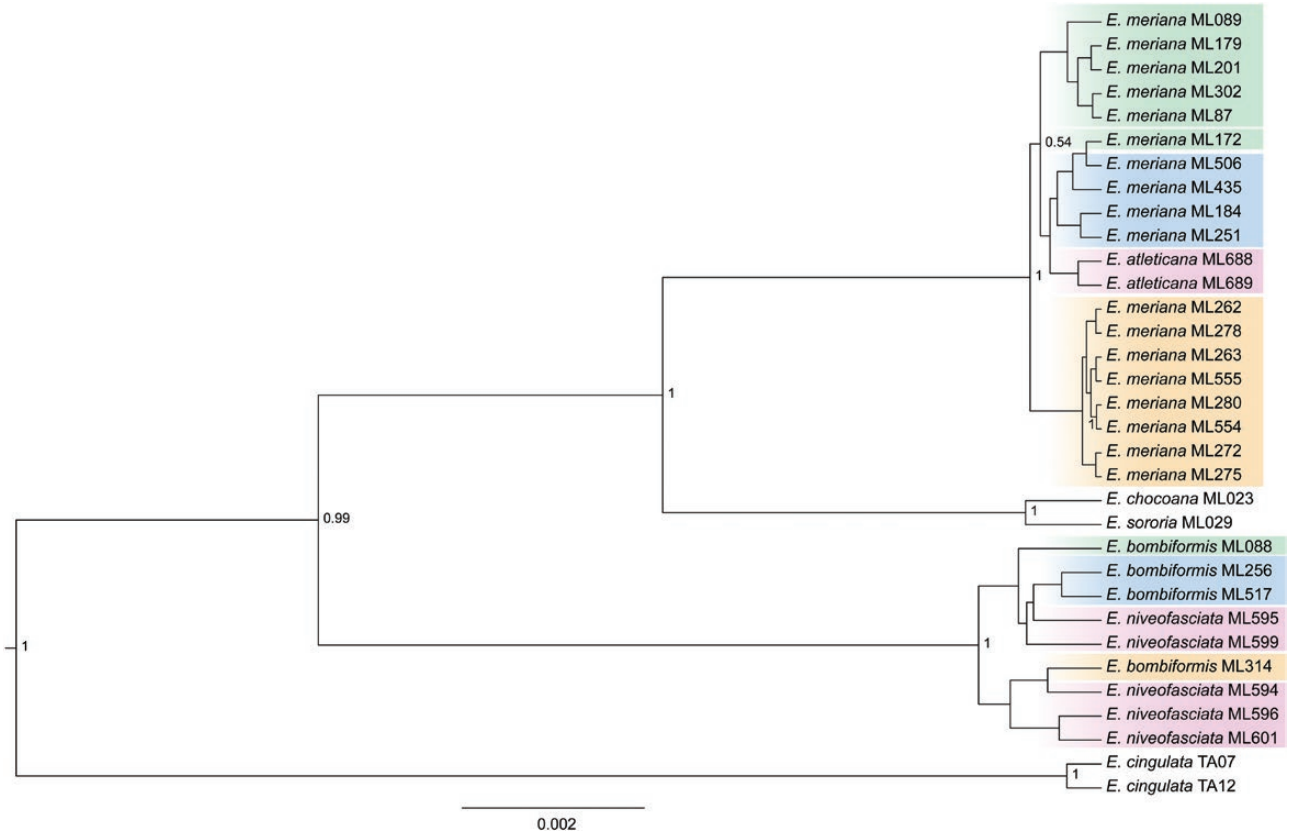

Fig. 3. Maximum clade credibility (MCC) species tree of the phylogenetic relationships in the *E. meriana* and *E. bombiformis* species complexes estimated under the multi-species coalescent model (MSC) using *BEAST in BEAST2. The species tree was estimated using the 50 most informative UCE loci from the 100% completeness dataset. Nodes without labels had posterior probabilities < 0.5. The colors of the clades correspond to geographic regions outlined in Fig. 1. Green: Central America, Orange: Choco region, Blue: Amazon Forest, and Pink: Atlantic Forest.

of lineages within the *E. meriana* and *E. bombiformis* complexes. Instead of clustering by color phenotype, both the mtDNA and UCE datasets recovered lineages that correspond to different biogeographic regions in the Neotropics: the Central American, Chocoan, Amazonian, and Atlantic tropical forests. These geographic areas are congruent with Pleistocene climatically stable areas previously identified by López-Uribe et al. (2014). Additionally, we found partial morphological differentiation of some of these lineages based on measurements of body size and the size of bands in tergites II and III of the metasoma. For the *E. meriana* complex, we found that *E. atleticana* individuals are morphologically distinct with narrower colored bands in T-II and T-III. For the *E. bombiformis* complex, we did not find any differentiation in the measurements for *E. niveofasciata* compared to the Amazonian lineage, but individuals from the Choco and Central American regions had slightly wider colored bands in T-II and T-III.

Taken altogether, our results suggest that different color phenotypes in these complexes represent color polymorphism within species rather than distinct species. It could be that color indicates early stages of divergence in some cases but for the most part independent lineages have not yet evolved (Moure 2000, Nemésio 2009). Dressler (1979) conducted a detailed examination of individuals of *E. meriana*, *E. bombiformis* and *E. seabrai* across their distributions to interpret color polymorphism in the context of divergence of these species. Before Dressler's work, these three species were very difficult to differentiate since taxonomists had been long using the metasoma coloration to delimit species. However, morphological characters on the hairs in the sternum V of males and the clypeo-orbital distance in both males and females are currently the most accurate and widely used characters to differentiate species (Dressler 1979, Roubik and Hanson 2004). Our results add to Dressler's interpretation that *E. bombiformis*, *E. seabrai* and *E. meriana* form Müllerian mimetic complexes across their distributions and that the frequency of the different color morphs might be driven by the relative abundance of co-mimics (Dressler 1979). In *Eulaema* bees, Müllerian mimicry is hypothesized to be the result of selection on a unique aposematic color morph that warns predators about the painful sting of *Eulaema* females. This warning coloration has likely driven one common phenotype in the Amazon region, which is dominated by red colors of *E. bombiformis* and *E. meriana*. In contrast, the yellow form is common west of the Andes in Colombia, Ecuador, and southern Central America but it is also found in northern Central America and the Atlantic Forest of Brazil which is referred to as the 'flavescens' pattern. Since the 'flavescens pattern' is present in the peripheral areas of distribution of *E. meriana* and not in the Amazon basin, Dressler hypothesized that this pattern might be relictual and was replaced by the 'bombiformis' red pattern in the Amazon due to the mimicry complexes.

Our results contribute to the growing literature describing color polymorphism likely driven by Müllerian mimicry. Examples include species of poison dart frogs (Comeault and Noonan 2011), wood tiger moths (Gordon et al. 2018), velvet ants (Wilson and Jahner 2015), and bumblebees. In a phylogenetic study of *Bombus trifasciatus*, Hines and Williams (2012) found that lineages within this species showed sub-lineages with several color patterns adhering to different mimetic groups, but that have no or very little

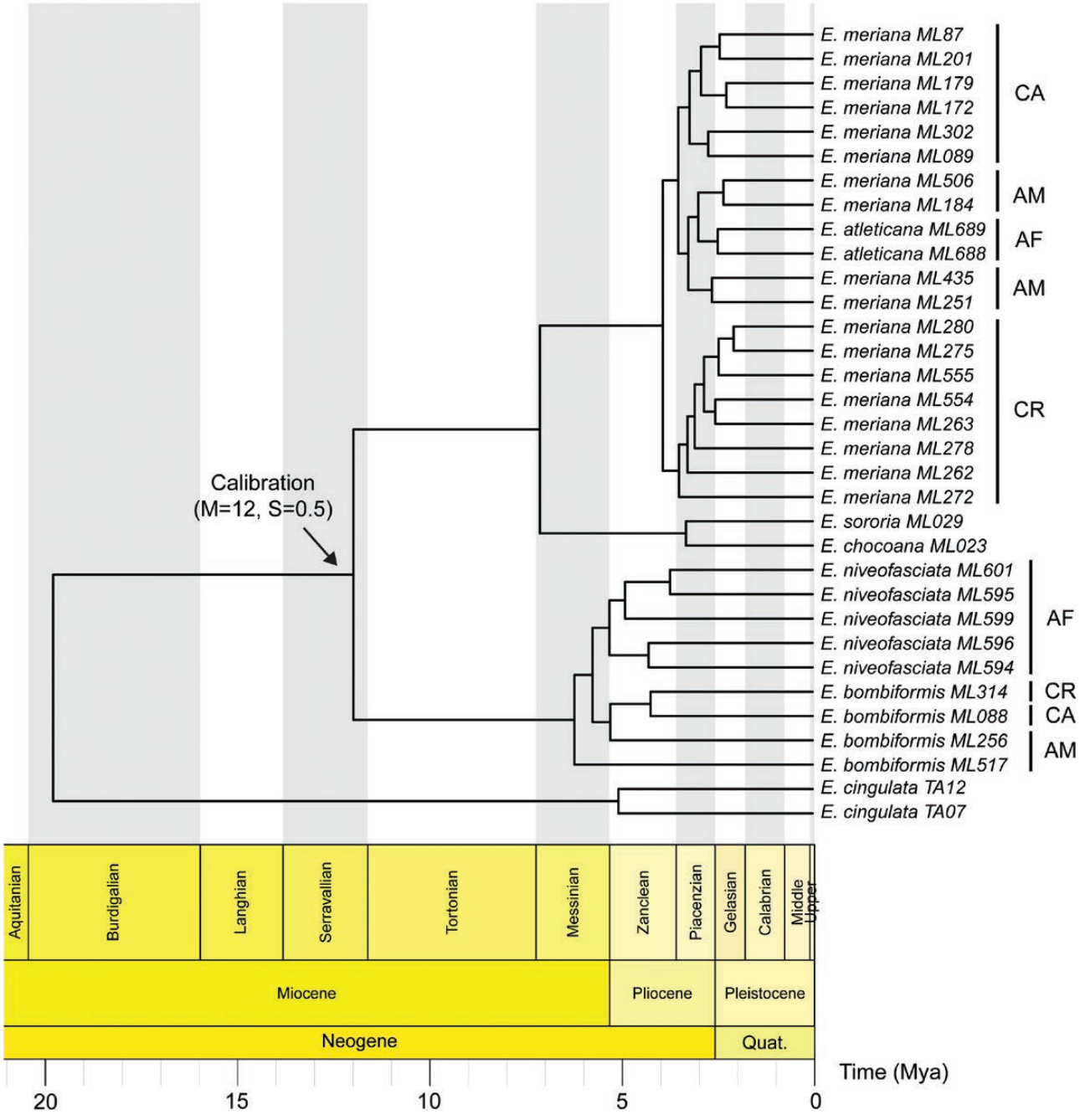

Fig. 4. Chronogram of the phylogenetic relationships in the *E. meriana* and *E. bombiformis* species complexes estimated using BEAST2 and 500 UCE loci. All nodes had a posterior probability of 1. The arrow indicates the node used for calibration of the tree and acronyms correspond to geographic regions outlined in Fig. 1. CA, Central America; CR, Choco Region; AM, Amazon Forest; AF, Atlantic Forest. In the calibration point, M = mean age, and S = confidence interval.

mitochondrial genetic variation. They also found intermediate color morphs that occur in neighboring areas of each sub-lineage. These intermediate color morphs in bumblebees are considered an indication of past gene flow between the neighboring populations or the result of later convergence due to local mimicry groups (Hines and Williams 2012). In our study, we identified an intermediate morph between the yellow ('flavescens') and the red ('bombiformis') patterns of *E. meriana* that occurs in the Central America and Choco regions. Dressler (1979) recognized a "brownish-yellow" morph that he included within the 'flavescens' pattern and that might refer to the orange individuals that we examined in this study.

Something important to note is that although we classified color polymorphism as a discrete character, the color variation is likely to be a continuum that varies from yellow to red, which is illustrated by the presence of the intermediate orange morph. We also found great variation in the tones of red, orange, and yellow exhibited by different specimens. Ezray et al. (2019) found that in some species of bumblebees the mimicry complexes exhibit continuous color variation instead of discrete patterns. This continuum is likely to result from transition zones where selection is relaxed, and mimicry is imperfect, making it subjective to score coloration as a discrete character. In our study, we did not sample individuals continuously across

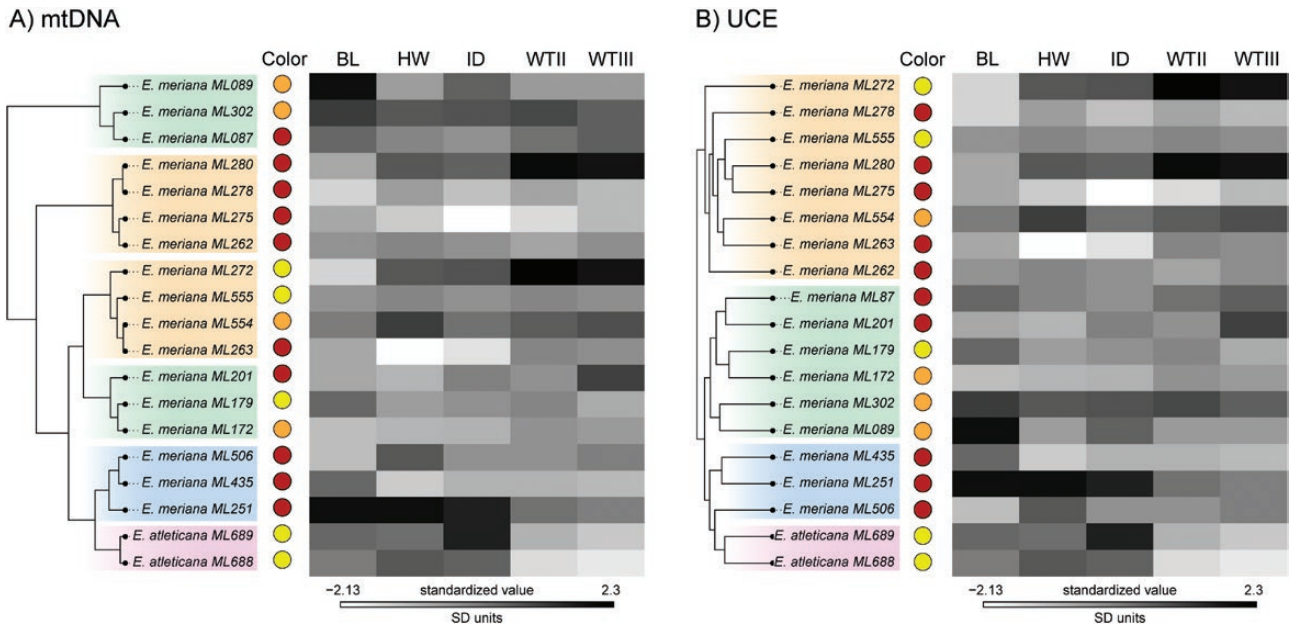

Fig. 5. Color phenotypes and phyloheatmap of the morphological characters of *E. meriana* and *E. atleticana*, visualized using the (A) mtDNA and (B) UCE phylogenies. Circles at the end of individual names indicate the color phenotype of that individual. Names above the phyloheatmaps indicate the character that was measured. BL, Body length; HW, Head width; ID, Intertegular distance; WTII, Width of colored bands on tergum II; WTIII, Width of colored bands on tergum III. In this phyloheatmap, each column of the measured characters was standardized to have the same variance prior to analysis. The scale below indicates how much each value deviates from the mean. The colors of the clades correspond to geographic regions outlined in Fig. 1. Green: Central America, Orange: Choco region, Blue: Amazon Forest, and Pink: Atlantic Forest.

the space and it is not possible to affirm that intermediate patterns are more abundant in transition zones which would require a larger sampling. A more detailed study of color patterns and frequencies in *Eulaema* and other species of orchid bees that are part of these mimetic complexes [e.g., *Eufriesea* spp., Dressler (1979), Roubik and Hanson (2004)] with methods such as machine learning (Ezray et al. 2019) could help us disentangle how mimicry complexes are formed and maintained in orchid bees.

The incongruencies we found between the mtDNA and UCE phylogenies highlight some of the limitations of using few genes in molecular datasets for species delimitation studies (Gueuning et al. 2020). In the previous *Eulaema* phylogeographic study by López-Uribe et al. (2014), the authors found population structure with the mitochondrial markers, but their microsatellite data did not show any differentiation between individuals in different geographic regions suggesting long-term male biased dispersal. Because male orchid bees fly long distances to collect aromatic compounds used in reproductive displays (Wikelski et al. 2010), this behavior may facilitate less genetic structuring in the nuclear compared to the mitochondrial markers. While the mitochondrial barcoding region is generally used as a marker for species identification and phylogeographic studies in insects (Packer et al. 2009), information from COI can both under and over-split species (Gibbs 2018). For example, in ants Prebus (2021) found that the inclusion of mitochondrial information in their datasets led to over-splitting of taxonomic units. Additionally, Gibbs (2018) found that COI barcodes of *Lassioglosum* (*Dialictus*) species (Hymenoptera: Halictidae) were only able to distinguish some species but in other cases, there was either under or over-estimation of species diversity within different species groups. Another potential driver of the incongruence between COI and genome-wide markers is that the effective population size of the mitochondrial genome is smaller than the nuclear genome. Thus, phylogeographic structure found with the mitochondrial data can result from a signature of recent gene flow that is not being captured by the nuclear UCE dataset (French et al. 2021). Nonetheless, using complete mitogenomes instead of only COI might be more informative and should be explored in future studies. Regardless of the causes of incongruence, our UCE dataset likely represents a more complete evolutionary history of the species under study, and it is a more robust dataset because it includes more loci and incorporates the maternal and paternal dynamics of the species evolutionary history.

Our estimated divergence times support the hypothesis that most of the lineage diversification in the highly polymorphic *E. meriana* occurred during the Pleistocene likely as a result of climatic instability associated with glaciations during this period. *Eulaema meriana* and *E. bombiformis* are primarily found in tropical rain forests and were likely susceptible to these climatic changes. Specifically, we found that divergence in the lineages of the *E. meriana* complex occurred sometime between ~5 and 2.5 Mya, similar to previous estimates based on mitochondrial evolutionary rates (López-Uribe et al. 2014). In the Pleistocene, the cycles of expansion and contraction of glaciers produced expansions and contractions of dry and wet environments which led to alternation of habitats. When the dry and cold weather expanded, the areas occupied by tropical and subtropical biota were reduced, leading to population differentiation (Hewitt 2004). Dressler (1979) hypothesized that the patterns of color variation in *Eulaema* might have originated in the Pleistocene due to these climatic changes. This might be possible considering our results but further investigation including a larger dataset with representatives from transition zones such as the Andean region would be necessary to test this hypothesis.

Despite the incongruence between the mitochondrial and nuclear data for the lineages with geographically restricted distributions, we found evidence for genetic differentiation of the Chocoan lineage from the rest of the *E. meriana* complex (Figs. 2B and 3).

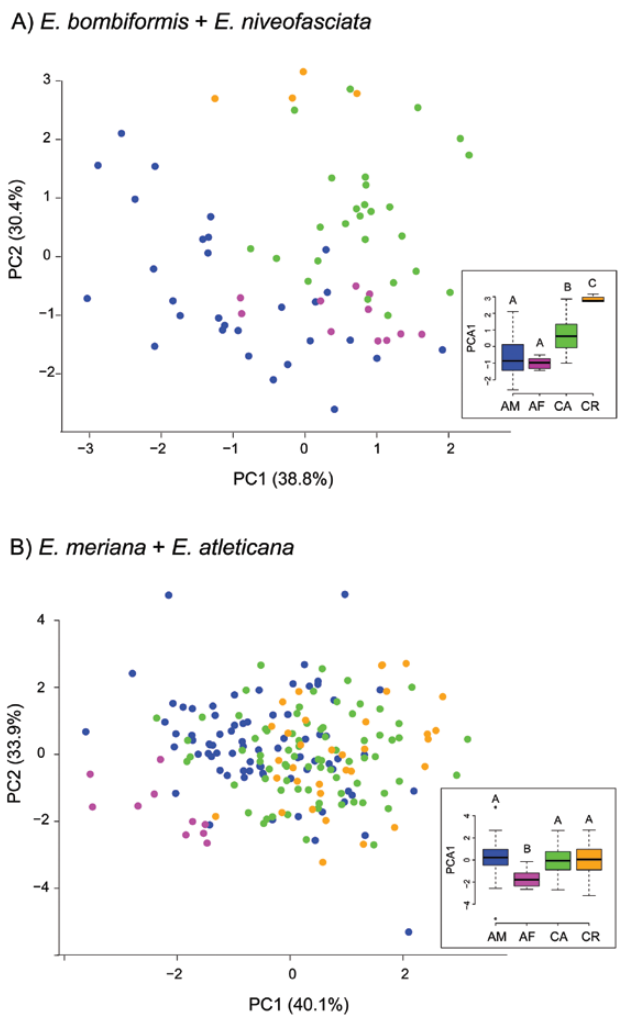

Fig. 6. Scatterplots of first against second principal component of the morphological measurements of the (A) *E. bombiformis* and (B) *E. meriana* complexes. Insets display boxplots of the first principal component between groups outlined by geographic regions: CA, Central America (Green); CR, Choco Region (Orange); AM, Amazon Forest (Blue); AF, Atlantic Forest (Pink). The letters above boxplots represent groups that are statistically differentiated after a Tukey's honest significant test. Colors represent individuals grouped by geographic regions.

Our dated phylogeny estimated that this lineage separated from the other populations sometime between ~6 and 2.5 Mya, which is congruent with the divergence of Chocoan and Amazonian populations as a result of a vicariant event due to the final uplift of the Andean mountains [~5 to 2.5 Mya, Hoorn et al. (2010)]. This would imply a later disconnection between the Choco and Central America while the Amazonian population has somehow maintained gene flow with the Central American lineage. Alternatively, the geological history of the Choco suggests that the modern tropical rainforest landscape was only established by the early Pleistocene [~2.7 Mya, Pérez-Escobar et al. (2019)] and that individuals from the Amazon Forest might have dispersed to the Choco after the Andean uplift without exchanging much gene flow with the Central American lineage. We did not find evidence of morphological differentiation of this lineage compared to the other populations of *E. meriana* but it could certainly represent a case of incipient speciation (Mayr 1942, De Queiroz 2007) in which the Choco lineage is in the process of diverging but there is still ongoing gene flow (Chou et al. 2021) and there are no clear color patterns or morphological characters that separate them (Ghisbain et al. 2020). To confirm whether there is an ongoing speciation process, a detailed sampling should be conducted in this region in the area between Ecuador (Choco Region) and Costa Rica (Central America).

For the *E. bombiformis* complex, we estimated older dates (7.5–3 Mya 95% HDP) which differs greatly from dates estimated by López-Uribe et al. (2014) using mitochondrial data. This could be an artifact of our sampling since we did not include a comprehensive number of individuals from all the geographic regions, or it could be an indicator of earlier diversification in this clade. The generation of more genomic data from individuals across the distribution of this species would help us disentangle the real time of diversification within this lineage. The species tree analysis did not show support for clades inside the *E. bombiformis* complex demonstrating incongruence between the genes and species trees on the structuration of these individuals. Although the mtDNA phylogeny shows *E. niveofasciata* as a clade separated from *E. bombiformis*, our ML and species tree analyses show no support for *E. niveofasciata* as a different lineage from *E. bombiformis*. The incongruence in the species tree might be due to a signature of gene flow between the different populations of *E. bombiformis* which might be missing from the IQ-TREE phylogeny due to incomplete lineage sorting. Our results for both the *E. meriana* and *E. bombiformis* complexes suggest either ancestral polymorphism or that there could have been recent gene flow between the Amazon and the Atlantic Forests populations. The latter has been confirmed for other taxa by ecological and palynological niche models that show historical connections during the Last Interglacial (LIG, 120 ka) and Last Glacial Maximum (LGM, 20 ka) that allowed migration (Ledo and Colli 2017). However, contemporary gene flow might be reduced due to the dry habitats that currently represent barriers between these two forests which might explain the difference in the mtDNA phylogeny (López-Uribe et al. 2014).

Taxonomic Recommendations

The phylogenetic and morphological analyses indicate that *E. atleticana* and *E. niveofasciata* represent Atlantic Forest populations of *E. meriana* and *E. bombiformis*, respectively, and not separate species. *Eulaema atleticana* was described relatively recently by Nemésio (2009) on the base of the bands of coloration in the metasoma being yellow but thinner than those of *E. flavescens*, and because this lineage is restricted to the Brazilian Atlantic Forest. This indicates that although we did not include samples from northern Venezuela, *E. flavescens* might also represent a lineage of *E. meriana* and not a different species since the yellow coloration pattern is present in other areas of the distribution and there are no other morphological characters that differentiate them. *Eulaema niveofasciata* was considered by Moure (2000) as a valid species but was later synonymized with *E. bombiformis* by Oliveira (2006). However, Nemésio (2009) reinstated the species status based on differences in coloration (pale-yellow/cream bands in the Atlantic individuals) and due to their presence being restricted to the Brazilian Atlantic Forest, similar to his justification to describe *E. atleticana*.

Given the unstable taxonomic history of *E. meriana* and *E. bombiformis* and the genetic, morphological and distribution data here presented, we propose to designate *E. atleticana* and *E. niveofasciata* as subspecies. The rank of subspecies represents a lower unit of biological organization that is relevant in conservation (Braby et al. 2012). Patten (2015) defined a subspecies as "a (morphologically) diagnosably distinct, geographically circumscribed

clade that does not form a distinct (neutral) genetic cluster or is not reciprocally monophyletic in relation to other such clades". Considering this definition and the recommendations by Phillimore and Owens (2006) and Braby et al. (2012) on better practices to designate subspecies, we think that *E. atleticana* and *E. niveofasciata* meet the requirements. Both lineages occur only in the Atlantic Forest which has suffered high levels of forest loss and fragmentation with only 14.5% of its original forested area remaining (Rosa et al. 2016). Although there have been historical connections between the Amazon and Atlantic Forests, currently gene flow is restricted by deforestation and the presence of dry habitats that separate them (Cerrado and Caatinga). Additionally, both lineages have been always recovered as a clade of closely related individuals in all our analyses and present morphological features that can differentiate them from other lineages. We then recognize the subspecies *Eulaema meriana atleticana* **comb. n.**, which can be identified by the yellow coloration of hairs on the metasoma, the configuration of these hairs into very thin bands, and by the pattern of hairs on sternite V of males typical of *E. meriana* (Dressler 1979, Nemésio 2009). We also recognize the subspecies *Eulaema bombiformis niveofasciata* **comb. n.**, which can be identified by the pale-yellow/cream coloration of hairs on the metasoma and by the pattern of hairs on sternite V of males typical of *E. bombiformis* (Dressler 1979, Moure 2000).

In general, our findings expand on the body of knowledge suggesting that coloration should not be used as a character for species identification in orchid bees. In *E. meriana* and *E. bombiformis*, other morphological characters such as the hairs on sternite V and the clypeo-orbital distance represent consistent characters for their identification (Dressler 1979). Additionally, the use of wing geometric morphometrics seems to be a promising approach for the morphological delimitation of cryptic or polymorphic species in mimicry complexes (Quezada-Euán et al. 2015). Our study illustrates the need for further examination of other species groups within *Eulaema* such as the *Eulaema seabrai* species complex, as well as several groups within the genus *Euglossa* in which delimitation is based on differences in integument coloration (Roubik and Hanson 2004). Additionally, future studies should consider a thorough taxonomic revision of the genus *Eulaema*, including specimens from all relevant biogeographic regions and detailed morphological studies that can clarify species status.

Supplementary Data

Supplementary data are available at *Insect Systematics and Diversity* online.

Acknowledgments

We would like to thank Christine Santiago for assistance in the molecular and morphological data collection, Rafael Ferrari for assistance with morphological measurements. We thank Nash Turley, Marcelo de Oliveira Gonzaga, André Nemésio, and Solange C. Augusto for the photos of bees in Fig. 1. We would also like to thank Shelby Kilpatrick and Tamires Andrade for discussions and assistance in the analysis of UCE data. Thanks to the López-Uribe Lab members at Penn State and Rodrigo Dias for comments on earlier versions of the manuscript. Specimen sampling was funded 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 MML-U). SSA was supported by a Fulbright-MinCiencias fellowship for doctoral studies in the United States. MML-U was funded through the USDA NIFA Appropriations under Projects PEN04716 Accession No. 102052 and the Lorenzo L. Langstroth Early Career Professor. MGB was funded through USDA project 2080-21000-019-000-D and NSF grant DEB-2127744. USDA is an equal opportunity provider and employer.

Author Contributions

Stephania Sandoval Arango (Conceptualization-Equal, Formal analysis-Lead, Investigation-Equal, Visualization-Lead, Writing – original draft-Lead, Writing – review & editing-Lead), Michael Branstetter (Data curation-Equal, Formal analysis-Equal, Resources-Equal, Writing – review & editing-Equal), Carolina Cardoso (Conceptualization-Equal, Data curation-Equal, Resources-Equal, Writing – review & editing-Equal), Margarita Lopez-Uribe (Conceptualization-Equal, Funding acquisition-Equal, Investigation-Equal, Project administration-Equal, Resources-Equal, Writing – review & editing-Equal)

Data Availability

Raw Illumina reads and contigs representing UCE loci have been deposited at the NCBI Sequence Read Archive (BioProject No. PRJNA859127). All extracted mtDNA sequences were previously deposited in GenBank by López-Uribe et al. (2014) and accession numbers are provided in Table S1. A complete list of relevant NCBI accession numbers for the UCE dataset are also available in Supplementary Table S1. Concatenated UCE matrices, mtDNA matrices, UCE alignments, tree files, additional analysis files, and the UCE bait sequence file have been deposited at Dryad (https://doi.org/10.5061/dryad.brv15dvd5). The PHYLUCE package and associated programs can be downloaded from github (https://github.com/faircloth-lab/phyluce). The bee-ant-specific baits used to enrich UCE loci can be purchased from Arbor Biosciences (https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-expert/ mybaits-expert-uce/). Data from this study are available from the Dryad Digital Repository: (Sandoval-Arango, 2023).

References Cited

Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, Yatomi J, Kajitani R, Okuno M, Yamaguchi K, *et al*. Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. Nat Commun. **2018**:9:1–13. https://doi.org/10.1038/s41467-018-06116-1.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham SON, Prjibelski AD, *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. **2012**:19(5):455–477. https://doi.org/10.1089/cmb.2012.0021.

Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Dawson A, Beltran SM, Perez-Espona S, Chamberlain N, Ferguson L, *et al*. Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. PLoS Genet. **2010**:6(2). https://doi.org/10.1371/journal.pgen.1000794.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. **2014**:30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, *et al*. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. **2019**:15(4):1–28. https://doi.org/10.1371/journal.pcbi.1006650.

Braby MF, Eastwood R, Murray N. The subspecies concept in butterflies: Has its application in taxonomy and conservation biology outlived its usefulness?. Biol J Linn Soc. **2012**:106(4):699–716. https://doi.org/10.1111/j.1095-8312.2012.01909.x.

Branstetter MG, Longino JT, Ward PS, Faircloth BC. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol Evol. **2017**:8:768–776. https://doi.org/10.1111/2041-210X.12742.

Branstetter MG, Müller A, Griswold TL, Orr MC, Zhu CD. Ultraconserved element phylogenomics and biogeography of the agriculturally important mason bee subgenus *Osmia* (*Osmia*). Syst Entomol. **2021**:46:453–472. https://doi.org/10.1111/syen.12470.

Brown KS. The biology of *Heliconius* and related genera. Annu Rev Entomol. **1981**:26(1):427–457. https://doi.org/10.1146/annurev.en.26.010181.002235.

Cameron SA. Phylogeny and biology of neotropical orchid bees (Euglossini). Annu Rev Entomol. **2004**:49(1):377–404. https://doi.org/10.1146/annurev.ento.49.072103.115855.

Chan KO, Grismer LL. To split or not to split? Multilocus phylogeny and molecular species delimitation of southeast Asian toads (Family: Bufonidae). BMC Evol Biol. **2019**:19:1–12. https://doi.org/10.1186/s12862-019-1422-3.

Chou MH, Tseng WZ, De Sang Y, Morgan B, De Vivo M, Kuan YH, Wang LJ, Chen WY, Huang JP. Incipient speciation and its impact on taxonomic decision: a case study using a sky island sister-species pair of stag beetles (Lucanidae: Lucanus). Biol J Linn Soc. **2021**:134(3):745–759. https://doi.org/10.1093/biolinnean/blab105.

Comeault AA, Noonan BP. Spatial variation in the fitness of divergent aposematic phenotypes of the poison frog, *Dendrobates tinctorius*. J Evol Biol. **2011**:24:1374–1379. https://doi.org/10.1111/j.1420-9101.2011.02258.x.

Danforth, B. N. 1999. Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial COI sequence data. Syst. Entomol. 24:377–393. https://doi.org/10.1046/j.1365-3113.1999.00087.x.

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more model, new heuristics and high-performance computing. Nat Methods. **2012**:9:7726–7772. https://doi.org/10.1038/nmeth.2109.

De Mendiburu F, Yaseen M. agricolae: Statistical procedures for agricultural research. R package version 1.4.0; **2020**. https://cran.r-project.org/package=agricolae.

De Queiroz K. Species concepts and species delimitation. Syst Biol. **2007**:56(6):879–886. https://doi.org/10.1080/10635150701701083.

Dressler R. *Eulaema bombiformis*, *E. meriana*, and Müllerian mimicry in related species (Hymenoptera: Apidae). Biotropica. **1979**:11(2):144–151. https://doi.org/10.2307/2387794.

Eltz T, Fritzsch F, Pech JR, Zimmermann Y, Ramírez SR, Quezada-Euan JJG, Bembé B. Characterization of the orchid bee *Euglossa viridissima* (Apidae: Euglossini) and a novel cryptic sibling species, by morphological, chemical, and genetic characters. Zool J Linn Soc. **2011**:163:1064–1076. https://doi.org/10.1111/j.1096-3642.2011.00740.x.

Ezray BD, Wham DC, Hill CE, Hines HM. Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum. Proc R Soc B Biol Sci. **2019**:286:20191501. https://doi.org/10.1098/rspb.2019.1501.

Faircloth BC. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. **2013**. http://dx.doi.org/10.6079/J9ILL.

Faircloth BC. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. **2016**:32(5):786–788. https://doi.org/10.1093/bioinformatics/btv646.

Ferrari BR, Melo GAR. Deceiving colors: recognition of color morphs as separate species in orchid bees is not supported by molecular evidence. Apidologie. **2014**:45:641–652. https://doi.org/10.1007/s13592-014-0280-7.

Freitas FV, Santos Júnior JE, Santos FR, Silveira FA. Species delimitation and sex associations in the bee genus *Thygater*, with the aid of molecular data, and the description of a new species. Apidologie. **2018**:49:484–496. https://doi.org/10.1007/s13592-018-0576-0.

French RLK, Bell AJ, Calladine KS, Acorn JH, Sperling FAH. Genomic distinctness despite shared color patterns among threatened populations of a tiger beetle. Conserv Genet. **2021**:22:873–888. https://doi.org/10.1007/s10592-021-01370-1.

Gautier M, Yamaguchi J, Foucaud J, Loiseau A, Ausset A, Facon B, Gschloessl B, Lagnel J, Loire E, Parrinello H, Severac D, Lopez-Roques C, Donnadieu C, Manno M, Berges H, Gharbi K, Lawson-Handley L, Zang LS, Vogel H, Estoup A, Prud'homme B. The genomic basis of color pattern polymorphism in the Harlequin Ladybird. Curr Biol. **2018**:28(20):3296–3302. e7. https://doi.org/10.1016/j.cub.2018.08.023.

Ghisbain G, Lozier JD, Rahman SR, Ezray BD, Tian L, Ulmer JM, Heraghty SD, Strange JP, Rasmont P, Hines HM. Substantial genetic divergence and lack of recent gene flow support cryptic speciation in a colour polymorphic bumble bee (*Bombus bifarius*) species complex. Syst Entomol. **2020**:45:635–652. https://doi.org/10.1111/syen.12419.

Gibbs J. DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome. **2018**:61(1):21–31. https://doi.org/10.1139/gen-2017-0096.

Gordon SP, Burdillat S, Mappes J. Phenotype-dependent mate choice and the influence of mixed-morph lineage on the reproductive success of a polymorphic and aposematic moth. Evol Ecol. **2018**:32:427–441. https://doi.org/10.1007/s10682-018-9944-5.

Grab H, Branstetter MG, Amon N, Urban-Mead KR, Park MG, Gibbs J, Blitzer EJ, Poveda K, Loeb G, Danforth BN. Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. **2019**:363:282–284. https://doi.org/10.1126/science.aat6016.

Grando C, Amon ND, Clough SJ, Guo N, Wei W, Azevedo P, López-Uribe MM, Zucchi MI. Two colors, one species: the case of *Melissodes nigroaenea* (Apidae: Eucerini), an important pollinator of cotton fields in Brazil. Sociobiology. **2018**:65(4):645–653. https://doi.org/10.13102/sociobiology.v65i4.3464.

Gueuning M, Frey JE, Praz C. Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. Mol Ecol. **2020**:29:4203–4220. https://doi.org/10.1111/mec.15629.

Guindon SG, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. **2010**:59(3):307–321. https://doi.org/10.1093/sysbio/syq010.

Harris RS. Improved pairwise alignment of genomic DNA. [Ph. D. dissertation]. Pennsylvania State University, University Park, Pennsylvania; **2007**. https://www.bx.psu.edu/~rsharris/rsharris_phd_thesis_2007.pdf.

Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data research article. Molecular. **2010**:27(3):570–580. https://doi.org/10.1093/molbev/msp274.

Hewitt GM. Genetic consequences of climatic oscillations in the quaternary. Philos Trans R Soc B Biol Sci. **2004**:359(1442):183–195. https://doi.org/10.1098%2Frstb.2003.1388.

Hines HM, Williams PH. Mimetic colour pattern evolution in the highly polymorphic *Bombus trifasciatus* (Hymenoptera: Apidae) species complex and its comimics. Zool J Linn Soc. **2012**:166(4):805–826. https://doi.org/10.1111/j.1096-3642.2012.00861.x.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. **2017**:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Hoorn C, Wesselingh FP, Ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, *et al.* Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science (80-). **2010**:330(6006):927–931. https://doi.org/10.1126/science.1194585.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability article fast track. Mol Biol Evol. **2013**:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kozak KM, Wahlberg N, Neild AFE, Dasmahapatra KK, Mallet J, Jiggins CD. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst Biol. **2015**:64(3):505–524. https://doi.org/10.1093/sysbio/syv007.

Lecocq T, Dellicour S, Michez D, Dehon M, Dewulf A, De Meulemeester T, Brasero N, Valterová I, Rasplus JY, Rasmont P. Methods for species delimitation in bumblebees (Hymenoptera, Apidae, *Bombus*): Towards an integrative approach. Zool Scr. **2015**:44:281–297. https://doi.org/10.1111/zsc.12107.

Ledo RMD, Colli GR. The historical connections between the Amazon and the Atlantic Forest revisited. J Biogeogr. **2017**:44:2551–2563. https://doi.org/10.1111/jbi.13049.

Lhomme P, Williams SD, Ghisbain G, Martinet B, Gérard M, Hines HM. Diversification pattern of the widespread holarctic cuckoo bumble bee, *Bombus flavidus* (Hymenoptera: Apidae): the east side story. Insect Syst Divers. **2020**:5(2):1–15. https://doi.org/10.1093/isd/ixab007.

Longino JT, Branstetter MG. Phylogenomic species delimitation, taxonomy, and 'Bird Guide' identification for the neotropical ant genus *Rasopone* (Hymenoptera: Formicidae). Insect Syst Divers. **2020**:4(2):1–33. https://doi.org/10.1093/isd%2Fixaa004.

López-Uribe MM, Zamudio KR, Cardoso CF, Danforth BN. Climate, physiological tolerance and sex-biased dispersal shape genetic structure of Neotropical orchid bees. Mol Ecol. **2014**:23:1874–1890. https://doi.org/10.1111/mec.12689.

Mayr E. Systematics and the origin of species. Columbia University Press, New York; **1942**.

Merrill RM, Dasmahapatra KK, Davey JW, Dell'Aglio DD, Hanly JJ, Huber B, Jiggins CD, Joron M, Kozak KM, Llaurens V, *et al*. The diversification of *Heliconius* butterflies: What have we learned in 150 years?. J Evol Biol. **2015**:28(8):1417–1438. https://doi.org/10.1111/jeb.12672.

Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop; **2010**. p. 1–8.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the Genomic Era. Mol Biol Evol. **2020**:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Moure JS. A checklist of the known euglossine bees. Atas do Simpósio sobre a Biota Amaz. **1967**:5:395–415.

Moure JS. The species of the genus *Eulaema* Lepeletier, 1841 (Hymenoptera, Apidae, Euglossinae). Acta Biol Parana. **2000**:29:1–70.

Muñoz-Ramírez CP, Bitton P-P, Doucet SM, Knowles LL. Mimics here and there, but not everywhere: Müllerian mimicry in *Ceroglossus* ground beetles?. Biol Lett. **2016**:12(9):20160429. https://doi.org/10.1098/rsbl.2016.0429.

Nemésio A. Orchid bees (Hymenoptera: Apidae) of the Brazilian Atlantic Forest. Zootaxa. **2009**:2041:1–242. https://doi.org/10.11646/zootaxa.2041.1.1.

Oliveira ML. Nova hipótese de relacionamento filogenético entre os gêneros de Euglossini e entre as espécies de *Eulaema Lepeletier*, 1841 (Hymenoptera: Apidae: Euglossini). Acta Amaz. **2006**:36(2):273–285. https://doi.org/10.1590/S0044-59672006000200018.

Packer L, Gibbs J, Sheffield CS, Hanner R. DNA barcoding and the mediocrity of morphology. Mol Ecol Resour. **2009**:9:42–50. https://doi.org/10.1111/j.1755-0998.2009.02631.x.

Padial JM, Miralles A, Riva ID, Vences M. The integrative future of taxonomy. Front Zool. **2010**:7:1–14. https://doi.org/10.1186/1742-9994-7-16.

Papadopoulou A, Anastasiou I, Vogler AP. Revisiting the insect mitochondrial molecular clock: the Mid-Aegean Trench Calibration. Mol Biol Evol. **2010**:27(7):1659–1672. https://doi.org/10.1093/molbev/msq051.

Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. **2004**:20(2):289–290. https://doi.org/10.1093/bioinformatics/btg412.

Patten MA. Subspecies and the philosophy of science. Auk. **2015**:132(2):481–485. https://doi.org/10.1642/AUK-15-1.1.

Pérez-Escobar OA, Lucas E, Jaramillo C, Monro A, Morris SK, Bogarín D, Greer D, Dodsworth S, Aguilar-Cano J, Sanchez Meseguer A, *et al*. The origin and diversification of the Hyperdiverse Flora in the Chocó Biogeographic Region. Front Plant Sci. **2019**:10:1–9. https://doi.org/10.3389/fpls.2019.01328.

Perry RK, Heraty JM. A tale of two setae: how morphology and ITS2 help delimit a cryptic species complex in Eulophidae (Hymenoptera: Chalcidoidea). Insect Syst Divers. **2019**:3(5):1–23. https://doi.org/10.1093/isd/ixz012.

Phillimore AB, Owens IPF. Are subspecies useful in evolutionary and conservation biology?. Proc R Soc B Biol Sci. **2006**:273(1590):1049–1053. https://doi.org/10.1098/rspb.2005.3425.

Plowright RC, Owen RE. The evolutionary significance of bumble bee color patterns: a mimetic interpretation. Evolution (N. Y). **1980**:34(4):622. https://doi.org/10.2307/2408017.

Prebus MM. Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): an integrative approach. Syst Entomol. **2021**:46:307–326. https://doi.org/10.1111/syen.12463.

Quezada-Euán JJG, Sheets HD, De Luna E, Eltz T. Identification of cryptic species and morphotypes in male *Euglossa:* morphometric analysis of forewings (Hymenoptera: Euglossini). Apidologie. **2015**:46:787–795. https://doi.org/10.1007/s13592-015-0369-7.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. **2018**:67(5):901–904. https://doi.org/10.1093/sysbio/syy032.

Ramírez SR, Roubik DW, Skov C, Pierce NE. Phylogeny, diversification patterns and historical biogeography of euglossine orchid bees (Hymenoptera: Apidae). Biol J Linn Soc. **2010**:100:552–572. https://doi.org/10.1111/j.1095-8312.2010.01440.x.

Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. **2012**:3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Rosa JF, Ramalho M, Arias MC. Functional connectivity and genetic diversity of *Eulaema atleticana* (Apidae, Euglossina) in the Brazilian Atlantic Forest Corridor: assessment of gene flow. Biotropica. **2016**:48:509–517. https://doi.org/10.1111/btp.12321.

Roubik DW, Hanson PE. Orchid bees of Tropical America: biology and field guide. Editorial INBio. **2004. p.** 370.

Sandoval-Arango S. Comparative phylogeography of three species of orchid bees (Apidae: Euglossini) with cross-Andean distributions. [M. S. thesis]. Universidade de São Paulo, Ribeirão Preto, Brazil; **2018**.

Sandoval-Arango S, Branstetter MG, Cardoso CF, López-Uribe MM. Data from: Phylogenomics reveals within species diversification but incongruence with color phenotypes in widespread orchid bees (Hymenoptera: Apidae: Euglossini). Dryad Digital Repository; **2023**. https://doi.org/10.5061/dryad.brv15dvd5.

Shevtsova E, Hansson C, Janzen DH, Kjærandsen J. Stable structural color patterns displayed on transparent insect wings. Proc Natl Acad Sci USA. **2011**:108(2):668–673. https://doi.org/10.1073/pnas.1017393108.

Sites JW, Marshall JC. Operational criteria for delimiting species. Annu Rev. **2004**:35:199–227. https://doi.org/10.1146/annurev.ecolsys.35.112202.130128.

Sousa-Paula LC, Pessoa FAC, Otranto D, Dantas-Torres F. Beyond taxonomy: species complexes in New World phlebotomine sand flies. Med Vet Entomol. **2021**:35:267–283. https://doi.org/10.1111/mve.12510.

Tagliacollo VA, Lanfear R. Estimating improved partitioning schemes for ultraconserved elements. Mol Biol Evol. **2018**:35(7):1798–1811. https://doi.org/10.1093/molbev/msy069.

Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. **2007**:56(4):564–577. https://doi.org/10.1080/10635150701472164.

Van Dam MH, Cabras AA, Lam AW. How the easter egg weevils got their spots: phylogenomics reveals Müllerian Mimicry in *Pachyrhynchus* (Coleoptera, Curculionidae). Syst Biol. **2022**:syac064. https://doi.org/10.1093/sysbio/syac064.

Wikelski M, Moxley J, Eaton-Mordas A, Lopez-Uribe MM, Holland R, Moskowitz D, Roubik DW, Kays R. Large-range movements of neotropical orchid bees observed via radio telemetry. PLoS One. **2010**:5(5):e10731 0–e107385. https://doi.org/10.1371/journal.pone.0010738.

Williams P. The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. Biol J Linn Soc. **2007**:92:97–118. https://doi.org/10.1111/j.1095-8312.2007.00878.x.

Williams JL, Zhang YM, LaPolla JS, Schultz TR, Lucky A. Phylogenomic delimitation of morphologically cryptic species in globetrotting *Nylanderia* (Hymenoptera: Formicidae) species complexes. Insect Syst Divers. **2022**:6(1):1–15. https://doi.org/10.1093/isd/ixab027.

Wilson JS, Jahner JP, Forister ML, Sheehan ES, Williams KA, Pitts JP. North American velvet ants form one of the world's largest known Müllerian mimicry complexes. Curr Biol. **2015**:25(16):R704–R706. https://doi.org/10.1016/j.cub.2015.06.053.

Zhang MY, Williams JL, Lucky A. Understanding UCEs: a comprehensive primer on using ultraconserved elements for arthropod phylogenomics. Insect Syst. Divers. **2019**:3(5):1–12. https://doi.org/10.1093/isd/ixz016.

Zhang YM, Sheikh SI, Ward AKG, Forbes AA, Prior KM, Stone GN, Gates MW, Egan SP, Zhang L, Davis C, *et al*. Delimiting the cryptic diversity and host preferences of *Sycophila* parasitoid wasps associated with oak galls using phylogenomic data. Mol Ecol. **2022**:31:4417–4433. https://doi.org/10.1111/mec.16582.