Genome-wide and cuticular hydrocarbon evidence shed light on potential drivers of speciation in a Neotropical ant species complex

Rubi Meza-Lázaro¹, Kenzy Peña-Carrillo², Chantal Poteaux³, Maria Lorenzi³, James Wetterer⁴, and Alejandro Zaldivar-Riveron¹

¹Universidad Nacional Autonoma de Mexico ²INIFAP ³Université Sorbonne Paris Nord ⁴Florida Atlantic University Harriet L Wilkes Honors College

October 14, 2021

Abstract

Reproductive isolation between geographically separated populations is generally considered the most common form of speciation. However, speciation may also occur in the absence of geographic barriers due phenotypic and genotypic factors such as chemical cue divergence, mating signal divergence and mitonuclear conflict. Here we performed an integrative study based on two genome-wide techniques, 3RAD and ultraconserved elements, coupled with cuticular hydrocarbon and mtDNA sequence data, to assess the species limits within the E. ruidum species-complex, a widespread and conspicuous group of Neotropical ants for which heteroplasmy has been recently discovered in some populations from southeast Mexico. Our analyses indicate the existence of at least five distinct species in this complex, two widely distributed along the Neotropics and three that are restricted to southeast Mexico and that apparently have high levels of heteroplasmy. We found that species boundaries in the complex did not coincide with geographic barriers. We therefore consider possible roles of alternative drivers that may have promoted the observed patterns of speciation, including mitonuclear incompatibility, cuticular hydrocarbon differentiation, and colony structure. Our study highlights the importance of simultaneously assessing different sources of evidence to disentangle the species limits of taxa with complicated evolutionary histories.

Genome-wide and cuticular hydrocarbon evidence shed light on potential drivers of speciation in a Neotropical ant species complex

Rubi N. Meza-Lázaro^{1*}, Kenzy I. Peña-Carrillo^{2,4,*}, Chantal Poteaux², Maria Cristina Lorenzi², James K. Wetterer³ and Alejandro Zaldívar-Riverón^{1**}

¹ Colección Nacional de Insectos, Instituto de Biología, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, C.P. 04510, Ciudad de México, México.

² Laboratoire d'Ethologie Expérimentale et Comparée, UR 4443, LEEC, Université Sorbonne Paris Nord, 99 avenue J.-B. Clément, F-93430 Villetaneuse, France.

³ Harriet L. Wilkes Honors College, Florida Atlantic University 5353 Parkside Drive, Jupiter, FL 33458

⁴ INIFAP, Campo Experimental General Terán, km 31 carretera Montemorelos-China, 67400, General Terán, N.L. Mexico.

*Contributed equally to the manuscript.

**Corresponding author: Alejandro Zaldívar-Riverón (azaldivar@ib.unam.mx).

Abstract

Reproductive isolation between geographically separated populations is generally considered the most common form of speciation. However, speciation may also occur in the absence of geographic barriers due to phenotypic and genotypic factors such as chemical cue divergence, mating signal divergence, and mitonuclear conflict. Here, we performed an integrative study based on two genome-wide techniques, 3RAD, and ultraconserved elements, coupled with cuticular hydrocarbon and mtDNA sequence data, to assess the species limits within the *E. ruidum*species-complex, a widespread and conspicuous group of Neotropical ants for which heteroplasmy has been recently discovered in some populations from southeast Mexico. Our analyses indicate the existence of at least five distinct species in this complex: two widely distributed across the Neotropics and three that are restricted to southeast Mexico and that apparently have high levels of heteroplasmy. We found that species boundaries in the complex did not coincide with geographic barriers. We therefore consider possible roles of alternative drivers that may have promoted the observed patterns of speciation, including mitonuclear incompatibility, cuticular hydrocarbon differentiation, and colony structure. Our study highlights the importance of simultaneously assessing different sources of evidence to disentangle the species limits of taxa with complicated evolutionary histories.

Introduction

Speciation in sexual organisms can be defined as the origin of reproductive isolation between two populations (de Queiroz 2007; Harrison and Larson 2014; Seehausen *et al.* 2014). Such reproductive isolation is thought to occur most often between geographically separated populations (Boomsma and Nash 2014), and allopatric speciation is thus considered the most common mode of species diversification (Coyne and Orr 2004). However, phylogenetic splits sometimes do not coincide with geographic dispersal barriers (Wollenberg Valero et al. 2019). In such cases, intrinsic factors may help to explain patterns of geographic distribution and species divergence (Seehausen et al. 2014).

Speciation can be driven by intrinsic barriers initiated by genetic incompatibilities, through genetic drift, as an indirect consequence of selection, or through genomic conflict (Presgraves 2010; Seehausen et al. 2014). One proposed mechanism initiated by genomic conflict is the evolution of incompatibilities in co-functioning mitochondrial (mt) and nuclear genes (Gershoni et al. 2009; Chou and Leu 2010; Burton and Barreto 2012; Crespi and Nosil 2013; Hill 2015). In eukaryotes, mitochondria are almost exclusively inherited through the maternal gamete (Birky 1995; Sato and Sato 2013; Radzvilavicius et al. 2017). Co-evolution of nuclear and mt genomes may result in inter-population hybrids that display cytoplasmic incompatibilities (Grun 1976), which create hybridization barriers that may contribute to speciation (Hill 2015, 2016, 2019). This hypothesis predicts that the mt genotype of each species will be functionally distinct, and that introgression of mt genomes will be prevented by mitonuclear incompatibilities that arise when heterospecific mt and nuclear genes attempt to co-function to enable aerobic respiration (Sloan 2015; Hill 2018, 2019).

Any process that promotes population divergence could also facilitate speciation (Pfennig et al. 2010). At the phenotypic level, communication signals play an important role in species identification and may contribute to or even drive reproductive isolation (Bradbury et al. 1998). In insects, cuticular hydrocarbons (CHCs) play a major role in species recognition, especially in social insects (Blomquist and Bagnères 2010; Sprenger and Menzel 2020). CHCs expressed on the insect's cuticle are essential to prevent desiccation, and have secondarily evolved in communication roles such as nestmate/non-nestmate discrimination, information about castes, and task specialization (Blomquist and Bagnères 2010; van Zweden and d'Ettorre 2010). CHCs can also be involved in interspecific discrimination (Bagnères and Wicker-Thomas 2010; Chung and Carroll 2015; Adams and Tsutsui 2020), and in ants they have been used as taxonomic tools to detect morphologically similar but chemically distinct lineages (e.g., Schlick-Steiner *et al.* 2010; Hartke *et al.* 2019; Lucas*et al.* 2002).

In the present study, we assessed the species boundaries within the *Ectatomma ruidum* species-complex, a widespread and conspicuous group of Neotropical ants, through integrative analyses of data from 3RAD, ultraconserved elements (UCEs), a fragment of the cytochrome oxidase 1 (cox1) mtDNA gene, and CHCs.

The ant genus *Ectatomma* Smith, 1858 (Ectatomminae) currently includes 15 valid species with mainly Neotropical distributions (Antweb 2021). Of these, *E. ruidum* (Roger) is perhaps the most widely distributed. This species was originally described from localities in Brazil, French Guiana, and Colombia (Roger 1860), though its type locality was subsequently restricted to Colombia (Kugler and Brown 1982). Currently, *E. ruidum* is known to occur from northern Mexico in the state of Tamaulipas to central Brazil, and also on some Caribbean islands (Aguilar-Velasco et al. 2016). This species inhabits a wide range of environments from sea level to 1600 m of elevation (Kugler and Brown 1982; Santamaría et al. 2009).

Ectatomma ruidum is of particular interest in evolutionary studies due to its taxonomic complexity, wide geographic distribution, and presence of heteroplasmy (occurrence of more than one mtDNA type within an individual) found in some of its populations. In the first phylogenetic study carried out for this speciescomplex, Aguilar-Velasco et al. (2016) proposed the existence of four evolutionary lineages plus a presumed hybrid population based on examination of external morphology and nuclear and mtDNA sequence data. Two of these lineages (*E.ruidum* spp. 1 and 2) have broad Neotropical distributions, whereas the others (*E. ruidum* spp. 3, 4, and 2x3) are apparently restricted to localities along the Pacific coast in southeast Mexico. The authors also reported considerable variation in the mt locus cox1, and the existence of nuclear mt paralogs (*numts*; Song *et al*. 2014) based on the presence of polymorphism in several chromatograms. Meza-Lázaro *et al.* (2018) subsequently assembled the mitogenomes of workers assigned to the putative species and the hybrid population proposed in the earlier study using NGS data. The mitogenomes of some of the populations from southeast Mexico (*E. ruidum* spp. 3, 4 and 2x3) had a high number of polymorphic sites, and a detailed examination indicated the presence of two functional mt genomes within these specimens.

Peña-Carrillo et al. (2021a) analyzed the CHC profiles of specimens assigned to *E. ruidum* from populations across its known geographic distribution, focusing on the heteroplasmic populations from southeast Mexico. The CHC profiles varied considerably among populations, supporting the existence of various evolutionary lineages within the complex. More recently, an acoustic study of the distress call for members of this complex found that a population in Huaxpaltepec, Oaxaca, Mexico, differs significantly in this trait from other examined populations, strongly suggesting that this population could represent a distinct, undescribed species (Peña-Carrillo et al. 2021b).

Here, we performed a comprehensive integrative systematic study based on different sources of molecular evidence to assess the species limits within the *E. ruidum* complex. Specifically, we coupled DNA sequence information generated using two reduced genome representation techniques, 3RAD (Bayona-Vásquez et al. 2019), and ultraconserved elements (UCEs; Faircloth et al. 2012; Faircloth 2017), with a large data set consisting of a mtDNA sequence fragment of a commonly barcoded locus (Hebert et al. 2003) and CHC profiles, with special attention to heteroplasmic populations from southeast Mexico.

Material and methods

Taxon sampling

We used worker specimens assigned to *E. ruidum* collected at localities across the Neotropics from central Mexico to Colombia, but with emphasis on the populations from the state of Oaxaca in southeast Mexico that contain high levels of heteroplasmy (Meza-Lázaro et al. 2018) (Figure 1). Our sample size varied by location, though in all datasets we included representative specimens of the four putative evolutionary lineages proposed in Aguilar-Velasco et al. (2016) study (*E. ruidum* spp. 1-4 and 2x3).

We examined a 626 bp fragment of the cox1 gene in 250 specimens assigned to *E. ruidum* and one specimen of *E. gibbum*, employing the latter as the outgroup. Of these sequences, 107 and 12 were obtained from Aguilar-Velasco et al. (2016) and Meza-Lázaro et al. (2018), respectively, whereas 132 were newly generated. We excluded from the data set all potential nuclear mt paralogous sequences that were detected based on the presence of internal stop codons or relationships that were evidently incorrect (Song et al. 2014). We also detected and excluded all fast-evolving secondary mt copies for the specimens assigned to *E. ruidum* spp. 3, 4 and 2x3, employing the phased cox1 sequences generated in Meza-Lázaro et al. (2018) as a reference. For the 3RAD and UCE data sets, we generated sequences for 35 (34 *E. ruidum*, one *E. tuberculatum* as outgroup) and 14 (13 of E. ruidum, one of E. gibbum as outgroup) specimens, respectively.

We also generated CHC profiles of 24 workers from two localities in Mexico (Puerto Morelos, Quintana Roo, 5 nests; Cuatode, Municipality of Santa María Tonameca, Oaxaca, 6 nests) and pooled them with the data set obtained in Peña-Carrilo et al. (2021a). The complete dataset comprised 132 workers and included two to five ants per colony. A list of the specimens used for the different data sets, their taxon assignments, localities information and DNA voucher and GenBank accession numbers is available in the supplementary material (Table S1).

DNA sequencing protocols and assembly procedures

All specimens were preserved in 96% ethanol until processed for DNA sequencing. We extracted genomic DNA from whole specimens using the EZ-10 Spin Column Genomic DNA Minipreps kit (BIO BASIC®), Toronto, Canada) and quantified it using the Qubit fluorometer system (High Sensitivity DNA kit, Life Technologies Inc., Carlsbad, CA, USA). We used 1:10 and 1:30 dilutions of DNA template for Sanger sequencing, and followed the procedures described by Aguilar-Velasco et al. (2016) for *cos1* amplification and sequencing. Sequences were edited and aligned based on their translated amino acids with the program Geneious version 10.1 (Biomatters, Ltd., Aukland, New Zealand).

We generated genome-wide sequence data using the 3RAD method (Bayona-Vásquez et al. 2019). This technique uses three restriction enzymes, two for the construction of dual-digest libraries and a third that cuts adapter-dimers formed by the phosphorylated adapter, thus increasing the efficiency of adapter ligation (Graham et al. 2015; Bayona-Vásquez et al. 2019). We digested 250 ng of the extracted genomic DNA for each sample using the XbaI and EcoRI-HF restriction enzymes (New England Biolabs; Beverly, MA, USA), which leave different sticky ends, and NheI (New England Biolabs; Beverly, MA, USA) to digest iTru adapter dimers. We ligated double-stranded iTru R1 and iTru R2.1 adapters onto each DNA fragment and ran a short PCR (13 -15 cycles) with the iTru5 and iTru7 primers obtained from Adapterama (Bayona-Vásquez et al. 2019). The resulting libraries were size selected in a 200-800 bp window and sequenced at the Genomic Sequencing Lab facilities at the University of California Berkeley. Libraries were sequenced using the 150 SRR HiSeq2500 Rapid, 10 pM, INDEX (124M Reads, 72% PhiX Aligned).

We used the process_radtags program implemented in the software pipeline Stacks version 2.0 (Catchen et al. 2011; Catchen et al. 2013) to demultiplex, clean, and trim the sequence data. We discarded any read with an uncalled base (-c) or with low quality scores (-q). We processed demultiplexed reads using the software pipeline ipyrad version 0.6.19 (Eaton 2014; Eaton and Overcast 2020) on the Miztli supercomputer owned by the Dirección General de Cómputo y de Tecnologías de Información y Comunicación, National Autonomous University of Mexico (DGTIC, UNAM). Reads from each sample were clustered using the program VSEARCH version 2.0.3 (https://github.com/torognes/vsearch) and aligned with the program MUSCLE version 3.8.31 (Edgar 2004).

To avoid the potential for false heterozygous calls due to clustering of paralogs (optimum clustering threshold; Eaton 2014), we followed the approach described by Ilut et al. (2014) to assess the level of sequence similarity at which two fragments are considered homologous. This approach minimizes the number of false homozygous (a single locus split into two) and false heterozygous (clustering of paralogs) loci in a clustering threshold series. We analyzed a clustering threshold series ranging from 0.80 to 0.98 in 0.02-0.03 increments. We also conducted maximum likelihood (ML) phylogenetic analyses as described below for the matrices derived from the above clustering threshold values to evaluate their level of nodal support. Based on the results obtained from these two approaches, we selected a clustering threshold value of 0.98 to build four matrices with min_samples_locus values of 25, 28, 30 and 33 including the outgroup and one with a min_sample_locus of 33 excluding it.

We generated UCE data from libraries following Branstetter et al. (2017). For this data set, we included previously generated UCE data from four male specimens (Meza-Lázaro et al. 2018). We fragmented up to 50 ng of input DNA to an average fragment distribution of 400-600 bp using a Qsonica Q800R (Qsonica LLC, Newton, CT) or a BioRuptor ($\hat{\mathbf{R}}$) Pico sonicator (Diagenode, Liége, Belgium). Following DNA fragmentation, we constructed sequencing libraries using the Kapa library preparation kit (Kapa Biosystems Inc., Wilmington, MA) and custom dual-indexing barcodes (Glenn et al. 2019). We purified PCR reactions 0.8 to 1.0X using Sera-Mag SpeedBeads (Thermo-Scientific, Waltham, MA, USA) (Rohland and Reich, 2012).

We pooled 10-12 libraries at equimolar concentrations for UCE enrichment, adjusting pool concentrations to 147 ng/ μ l. We used a total of 500 ng of DNA (3.4 μ l each pool) for each enrichment. We enriched each pool using the bait set 'ant-specific hym-v2' (Branstetter et al. 2017), which has 9,446 custom-designed probes (MyBaits, MYcroarray, Inc., Ann Arbor, MI, USA) targeting 2,524 UCE loci and 452 baits targeting 16 commonly sequenced exons. The enriched library quality was verified using an Agilent TapeStation 2200 (Agilent Tech, Santa Clara, CA, USA). We sent pools to the University of Utah Genomics Core facility and to the Georgia Genomics Facility at the University of Georgia, where they were sequenced on an Illumina HiSeq 2500 (PE150) and an Illumina NextSeq v2300 (PE150), respectively.

Raw data were demultiplexed and converted from BCl to FASTQ by the sequencing facilities. We used the program PHYLUCE version 1.5.0 and its associated programs (Faircloth 2016) for assembly and alignment of the UCE data. We cleaned and trimmed raw reads using ILLUMIPROCESSOR (Faircloth 2013). The cleaned and trimmed reads were assembled de novo using the program ABySS version 1.3.6 (Simpson et al. 2009). We mapped the assembled contigs to the hym-v2 bait database to identify individual UCE loci, to remove paralogs and to generate a list of shared UCE loci. We sorted out data by locus and aligned each one with the program MAFFT version 7.130b (Katoh et al. 2019). The resulting alignments were filtered and trimmed with the program Gblocks version 0.91b (Castresana 2000; Talavera and Castresana 2007). We analyzed matrices with 75, 80, 90, 95 and 100% taxon occupancy (percent of taxa required to be present at each locus).

We also followed the Tutorial II: Phasing UCE data (Faircloth 2021) to call for SNPs, which is derived from the procedure described by Andermann et al. (2019). The above workflow requires an individual-specific "reference" that can be aligned against raw reads. Hymenopteran males are haploids, and thus we only expected homozygous loci for them. We used edge-trimmed exploded alignments as reference contigs and aligned raw reads to them. We exploded the edge trimmed alignments to create separate FASTA files for each sample using phyluce_align_explode_alignments. We used BWA-MEM to map the FASTQ read files against the contig reference database for each sample. We sorted the reads within each bam file into two separate bam files using phyluce_snp_phase_uces. We built three final matrices based on filtering UCE loci with 85%, 90% and 100% taxon occupancy. We also used the exploded alignment and raw reads of the 100% taxon occupancy matrix to build an additional matrix phasing the data and calling a single variant SNP per locus.

Phylogenetic analyses

Cox1. We conducted a maximum likelihood (ML) phylogenetic analysis for the cox1 data set using the program IQTREE with 10,000 ultrafast bootstrap replicates (UFBoot2; Nguyen et al. 2015; Hoang et al. 2018) at the CEBIV, Austria web server (Trifinopoulos et al. 2016). We considered three partitions according to codon positions and selected the most appropriate models of evolution for each partition with the program ModelFinder (Kalyaanamoorthy et al. 2017). The selected evolutionary models were position 1 = HKY+F+I, position 2 = HKY+F, position 3 = TIM2+F+G4. Branches with UFBoot2 support [?]95% were considered as well supported (Hoang et al. 2018). We also built a haplotype network in POPART (Leigh and Bryant 2015) using TCS (Clement et al. 1997) to have a better visualization of the relationships among cox1 haplotypes.

3RAD. We carried out ML and Bayesian phylogenetic analyses with the four selected matrices including the outgroup (clustering threshold = 98%, min sample locus= 25, 28, 30, 33). The ML analyses were conducted with the program RAxML version 8.0 (Stamatakis 2014) using the GTR-GAMMA model. Branch support was estimated using the automatic bootstrap function, which calculates a stopping rule to determine when enough replicates have been generated (Pattengale et al. 2010). We conducted the Bayesian analyses with the program ExaBayes version 1.5 (Aberer et al. 2014). These analyses were run using the generalized time-

reversible model (GTR+G) and five independent MCMC chains of 1,000,000 generations each. The first 100,000 trees (10%) were discarded as burn-in for each MCMC run prior convergence (i.e. when maximum discrepancies across chains <0.1). We assessed burn-in, convergence among runs and run performance examining the resulting parameter files with the program TRACER version 1.7.0 (Rambaut et al 2018). We computed consensus trees using the consensus utility of ExaBayes.

We also employed the program SVDquartets version 1.0 (Chifman and Kubatko 2014) implemented in PAUP version 4.0a (Swofford et al. 2003) to carry out lineage-level phylogenetic analyses with SNP data using the above matrices. SVDquartets is a fast and robust method implemented in the latter program for species- and lineage-level phylogenetic analysis using multi-locus or SNP data under the coalescent model. We assessed variability in the estimated tree using a nonparametric bootstrapping with 500 replicates.

UCEs. We conducted ML analyses for the five UCE matrices (75, 80, 90, 95, 100% taxon occupancy) using the program RAxML version 8 (Stamatakis 2014) with the best tree plus rapid bootstrap search ('-f a' option) and 200 bootstrap replicates. We used the GTR+ Γ model of sequence evolution for the best tree and bootstrap searches. We carried out these analyses using three different partition schemes (unpartitioned, data partitioned by locus, data pre-partitioned by locus). We selected the best evolutionary model for these partitions using the program PARTITIONFINDER version 2 (Lanfear et al. 2017) based on the Bayesian Information Criterion (BIC) and the reluster option, which is more appropriate for larger data sets.

We carried out Bayesian analyses with the program Exabayes version 1.4.1 (Aberer et al. 2014). Each analysis consisted of two independent runs of 10 million generations each, two independent runs with Metropolis-Coupling in parallel to better sample parameter space, three heated and one cold chain per run, and sampling trees every 1000 generations. We linked branch lengths across partitions and ran each partitioned search for one million generations. Mixing and stationarity were monitored with the program TRACER version 1.6 (Rambaut and Drummond 2003). We built consensus trees using the consensus utility contained in the program Exabayes version 1.4.1 (Aberer et al. 2014) using a burn-in of 25%. To evaluate for the presence of reticulation, we used the neighbor-net method (Bryant and Moulton 2004) implemented in the program SplitsTree (Huson and Bryant 2006) for computing an unrooted phylogenetic network based on the alignment of the concatenated phased loci with 100% taxon occupancy.

For tree estimation based on the multispecies coalescent model, we generated gene trees from the 100% taxon occupancy matrix, phasing loci with the program RaxML with 200 bootstrap replicates. We then used the resulting gene trees to carry out analyses with the program ASTRAL v.4.10.8 (Mirarab et al. 2014; Mirarab and Warnow 2015), which used unrooted trees and missing data. We calculated nodal support with 200 multi-locus bootstrap replicates (Seo 2008).

Genetic structure and species delineation analyses

We employed the program STRUCTURE version 2.3.4 (Pritchard et al. 2000) implemented in the ipyrad.analysis toolkit (https://ipyrad.readthedocs.io/analysis.html), to assess patterns of genetic structure and admixture among the examined populations using the 3RAD dataset. We used the 98_33 matrix (clustering threshold=98 and min_sample_locus = 33, outgroup excluded) to perform an individual-based Bayesian clustering analysis. We used an admixture model with correlated frequencies and assessed values of population differentiation (K) in 15 independent runs for each K from 2 to 8. All runs were conducted with the first 250,000 being discarded as burn-in. Figures were generated based on the iterations with the highest posterior probability. The optimal K value was determined based on the highest average likelihood value [LnP(D)] obtained (Evanno et al. 2005).

We carried out species delimitation analyses with the 3RAD data using the program Bayesian Phylogenetics & Phylogeography version 3.3 (BPP; Yang and Rannala 2010; Yang and Rannala 2014; Flouri et al. 2018). BPP evaluates speciation models using a reversal jump Markov Chain Monte Carlo (rjMCMC) algorithm to determine whether to collapse or retain nodes in the phylogeny, assuming no admixture following a speciation event (Yang and Rannala 2010). BPP requires an input guide tree representing the species phylogeny with all possible species (Leaché and Fujita 2010). We used the topology derived from the analysis with the 98_33

matrix as guide tree, employing two different species assignment criteria: 1) a morphospecies assignment following Aguilar-Velasco et al. (2016), and 2) based on geographically congruent clades containing at least two samples.

We randomly subsampled the 98_33 matrix to produce three different matrices with 200 loci and 10 with 50 loci. We examined three sets of parameters varying the ancestral population size (ϑ) and root age (τ). The first assumed large ancestral population sizes and deep divergences [ϑ G (1, 10) and τ 0G(1, 10)], the second small ancestral population sizes and shallow divergences among species [ϑ G(2, 2000) and τ 0G(2, 2000) 3], and the third large ancestral populations sizes and relatively shallow divergences among species [ϑ G(1, 10) τ 0 G(2, 2000)]. We ran analyses for all models without data to separately evaluate the effects of the data and parameters. We subsequently ran seven replicates for one of the matrices with 50 loci with each set of parameters to check whether the analyses were run long enough, and then ran analyses for all matrices with five replicates each starting from different seeds. A significant posterior probability (PP [?] 0.95) value was employed across all runs to retain a given node (i.e., indicating lineage splitting). All analyses were run for 500,000 generations (first 10,000 were burn-in), with a sampling interval of 50.

We conducted a Bayes Factor species delimitation (BFD) analysis (Grummer et al. 2014; Leaché et al. 2014) for the phased UCE 100% taxon occupancy matrix calling a single variant SNP per locus. The BFD approach compares candidate species delimitation models with different numbers of species, estimating the marginal likelihood of each competing species delimitation model, ranking models by marginal likelihood and estimating Bayes factors to assess support for model rankings (Kass and Raftery 1995).

We also conducted a BDF analysis following Leaché and Bouckaert's (2018) tutorial. This tutorial implements SNAPP (Bryant et al. 2012), available in the BEAST2 platform (Bouckaert et al. 2014). SNAPP bypass the necessity of having to explicitly integrate or sample gene trees at each locus, and it codes SNP data as follows: individual homozygous for the original state= "0," heterozygous= "1," homozygous for derived state = "2". We used a python function to extract biallelic SNPs directly from allele Multiple Sequence Alignments (MSAs) (snps_from_uce_alignments.py, available from: github.com/tobiashofmann88/snp_extraction_from_alignments/; Andermann et al. 2019). We compared species delimitation models differing in the number of species. The base model had five species (*E. ruidum*spp. 1-4, 2x3), whereas the alternative ones were collapsed into one, two, three and four species in different combinations. The path sampling parameters, which are used to estimate marginal likelihood, were set to 12 and 24 steps with chainLength = 1,000,000.

All the analyses conducted using the 3RAD and UCE data sets were run on the CIPRES Web Portal (Miller et al. 2010) or on the the Miztli supercomputer owned by the Dirección General de Cómputo y de Tecnologías de Información y Comunicación, Universidad Nacional Autónoma de México (DGTIC, UNAM).

Genetic distances

We calculated uncorrected genetic distances for the 3RAD (98₋ 33 matrix), UCEs and *cox1* data sets. *Cox1* distance matrices were calculated using average distance per population, whereas for the 3RAD data set distances were calculated only using SNPs. All genetic distances were calculated using the program Mega X (Kumar et al. 2018).

Chemical distances based on CHC analyses

New CHC extractions were performed following Peña-Carrillo et al. (2021a). We used the Bray-Curtis distances as dissimilarity measures to calculate chemical distances between the 156 examined specimens based on the relative abundances of the simplified chemical profile. This was obtained by summing the percentages of hydrocarbons with the same carbon-chain length by class (Peña-Carrillo et al. 2021a). We then calculated the chemical distance between populations (centroids) and used them to obtain a hierarchical cluster dendrogram using the program Primer 6 & Permanova + (Anderson 2017). We also performed Mantel tests based on Pearson's product-moment correlation, running analyses with a maximum of 999 permutations to correlate chemical, genetic and geographic distances among the examined populations. Mantel tests were performed with the Vegan package (Oksanen et al. 2015) in R Studio (Version 3.5.2).

Results

Genome-wide data

The samples included in the 3RAD data set had from 75193 to 1007069 reads, and the generated matrices contained 986–7094 loci (Table S2). The outgroup employed for this data set, *E. tuberculatum*, shared between 375 and 1322 loci with the ingroup. The matrices derived from the unphased UCE data varied from 642 to 2196 loci and from 508859 to 1,817,455 characters (100% and 75% taxon occupancy matrices, respectively; Table S3). The phased UCE matrices had considerably fewer loci than those with unphased data (737 and 390 loci for the 85 and 100% taxon occupancy matrices, respectively), though the unphased data contained more parsimoniously informative sites (Table S3). SNP codification as a binary matrix for the UCE data showed the presence of heterozygous loci in the male samples which are also present in the nucleotide alignments.

Phylogenetic analyses

The Bayesian phylogram derived from the cox1 data set (Figure S1) recovered the putative species *E. ruidum* sp. 1 and *E. ruidum* sp. 2 each as monophyletic (PP= 1.0 in both cases). The specimens assigned to *E. ruidum* sp. 3–4 and 2x3 on the other hand appeared intermingled in a single clade (PP= 1.0). *Ectatomma ruidum* sp. 1 was sister to the remaining taxa, but with low support (PP= 0.87). Most of the internal relationships within *E. ruidum*sp. 1, formed by specimens from southern Mexico, Guatemala, Honduras, Nicaragua, Costa Rica, Colombia, Venezuela and the Lesser Antilles, were unresolved. In contrast, *E. ruidum* sp. 2 had an evident geographic structure, being composed of four main, significantly-supported subclades. One of these clades contained specimens from Ecuador (PP= 1.0), a second specimens from the remaining South and Central American localities (PP = 0.87), a third specimens from Quintana Roo in southeast Mexico (PP = 0.99), and the fourth contained specimens from southeast, central, and northeast Mexico (PP= 0.66).

The cox1 haplotype network showed that the *E. ruidum* haplotypes are grouped into three main haplogroups, which are separated from each other by 17 to 20 mutational steps (Figure 2D). Two of these haplogroups are each represented by specimens assigned to *E. ruidum* spp. 1 and 2, whereas the remaining one had members of *E. ruidum* spp. 3, 4 and 2x3. Moreover, the haplogroup with specimens of *E. ruidum* sp. 2 was divided into various geographically structured clusters.

The phylogenetic analyses with the 3RAD data set yielded well resolved, highly supported topologies. All the topologies derived from the ML, Exabayes and SVDquartets' analyses recovered four main clades with significant support (Figures 2B, S2). One of these clades contained members of E. ruidum sp. 1, which was sister to a second clade represented by the specimens of E. ruidum sp. 2. These two taxa were sister to the two remaining main clades, one of which contained four of the five specimens of E. ruidum sp. 3 and the two specimens of E. ruidum sp. 4, and the remaining one the specimen assigned to E. ruidum sp. 2x3 from a locality near Pinotepa Nacional, Oaxaca, and the specimen of E. ruidum sp. 3 from Guerrero, Mexico. The main clade formed by samples assigned to E. ruidum sp. 2 was further divided into two subclades. One was exclusively composed of specimens distributed from southeast (Yucatan and Chiapas) to northeast (Tamaulipas) Mexico (E. ruidum sp. 2A), whereas the second was comprised of specimens from Central America, including Quintana Roo in southeast Mexico, and South America (E. ruidum sp. 2B).

Most of the UCE analyses carried out with the ML, Exabayes and species tree methods recovered the above four clades with the same relationships among them mostly with strong support (Figure 2C, Supplementary Material S2). However, in contrast to most of the 3RAD topologies, the ML phylograms based on the 90, 95 and 100% occupancy matrices recovered *E. ruidum* sp. 1 as sister to the clade formed by *E. ruidum* sp. 3, 4 and 2x3 (BTP= 100).

Genetic structure

The STRUCTURE analyses carried out with the 3RAD data set displayed the highest value of Ln P(D) at K = 7, though the Evanno method had its highest value at K = 3 (Figure 3). At K = 2, we recovered one cluster with members of E. *ruidum* sp. 1 and E. *ruidum* sp. 2 and another with those of E. *ruidum*

spp. 3, 4 and 2x3. At K=3 the specimens of *E. ruidum* sp. 1 from Guatemala and Trinidad and Tobago formed a relatively well-differentiated cluster. At subsequent K values, the specimens from Pinotepa and localities in Guerrero in Mexico, and the ones from Trinidad and Tobago were each recovered as exclusive clusters, whereas at K=7 and 8 the specimens representing *E. ruidum* sp. 3 (except the one from Guerrero), *E. ruidum* sp. 4, *E. ruidum* sp. 2 from Quintana Roo, Mexico and Colombia, and the one from Guatemala each formed well-differentiated clusters.

$Species\ delineation$

The BPP analyses conducted using the 3RAD data set significantly supported the species model that considered five evolutionary lineages (PP > 0.95; Table S3), the four putative species proposed by Aguilar Velasco et al. (2016), *E. ruidum* spp. 1-4, and the specimens from Pinotepa and the state of Guerrero that these authors regarded as *E. ruidum* sp. 2x3 and as part of *E. ruidum* sp. 3, respectively. In contrast, when we assigned the samples according to their geographic proximity, our results widely varied (Table S4).

The Bayes factor comparisons of the two tested species delimitation models using the phased UCE data consistently favored a five species model, with E. ruidum spp. 1, 2, 3 and 4, as well as the specimens from Pinotepa and Guerrero, Mexico, each representing separate species (Table S4).

Chemical distances based on CHC analyses

The cluster analysis carried out on the CHCs yielded two main clusters (Figure 2A). One belonged to the hydrocarbon profiles of *E. ruidum* sp. 1 and is clearly separated from the second, which contained two subclusters, one with the *E. ruidum* sp. 2 samples and the second with specimens assigned to *E. ruidum* sp. 3 and 4. The*E. ruidum* sp. 2 cluster showed that the CHC profiles of the populations from Colombia and Puerto Morelos, in southeast Mexico, were more similar to each other than those of the population from Coyula, Oaxaca, Mexico (Figure 2A). We also found that the CHC profiles of E. *ruidum* sp. 3 were different from those of the populations of *E. ruidum* sp. 4.

Chemical distances were significantly correlated with the genetic distances obtained for the 3RAD (Mantel test, r = 0.855, P = 0.001), UCE (r = 0.876, P = 0.016), and cox1 (r = 0.847, P = 0.004) data sets (Table S6; Figure 4). In contrast, there was no significant correlation between geographic and chemical distances (r = -0.022, P = 0.354), nor between geographic and genetic distances (Table S6) calculated from 3RAD (r = 0.322, P = 0.141), UCEs (r = 0.052, P = 0.225), cox1 primary haplotypes (r = 0.217, P = 0.205) and cox1 secondary haplotypes (r = -0.258, P = 0.826).

Discussion

Here we used simultaneous assessment of nuclear and mtDNA along with CHC data to disentangle the species limits in a morphologically conserved insect taxon with an intricate evolutionary history. Morphological and genetic (mainly mt) evidence gathered in previous studies suggested the probable existence of four distinct evolutionary lineages within the *E. ruidum* complex (Aguilar-Velasco et al. 2016; Meza-Lázaro et al. 2018). However, those studies were inconclusive due to the presence of mt heteroplasmy in individuals of some populations from Oaxaca, which led to the reconstruction of clades with considerably long branches due to the preferential sequencing of the mt haplotype with the faster substitution rate over the alternative haplotype from the same specimen. Moreover, the mt markers recovered highly genetically structured populations, which is frequent in social hymenopterans due to extreme female philopatry (Johnstone et al. 2012; Hakala et al. 2019), whereas the nuclear markers had almost no variation. Below we discuss our results from two different nuclear genomic sequence data, which, together with the mt and CHC information, overcome limitations of the previous works, providing a robust framework of species delimitation that can be employed for further taxonomic and evolutionary studies in the group.

Integrative species delimitation

Integrating molecular and phenotypic data using different methodologies is widely assumed to be the most effective approach for delimiting species (Dayrat 2005; Will et al. 2005; Schlick-Steiner et al. 2010; Pante et

al. 2015). Integrative taxonomy provides statistical rigor for species delineation and validation of species, as well as for assignment of specimens to a given species group, improving the detection of cryptic diversity and the inference of relationships among species (Leaché and Fujita 2010; Schlick-Steiner et al. 2010; Edwards and Knowles 2014; Leavitt et al. 2015). This approach allows the identification of concordant patterns of divergence based on different sources of information (e.g. unlinked genetic loci, morphology, behavior), thus revealing cases of full lineage separation, since it is highly unlikely that a coherent pattern of character agreement emerges by chance (Padial et al. 2010).

Our molecular-based analyses using the cox1, 3RAD and UCEs data sets, together with analysis of CHC variation, yielded strongly congruent evidence for the existence of five different evolutionary lineages among the examined populations of *E. ruidum*. Our results support the existence of the four species suggested by Aguilar-Velasco et al. (2016) (*E. ruidum* spp. 1-4) and also a fifth species corresponding to what they suggested to be a hybrid population (*E. ruidum* 2x3). Four of these species were found in localities situated along the lowlands of the Sierra Madre del Sur in Oaxaca and Guerrero, Mexico. Moreover, the possibility of a sixth species in the area is plausible according to recent chemical, genetic and acoustic studies (Peña-Carrillo et al. 2021a, b).

We consistently recovered the specimens assigned to E. ruidum sp. 1 as a well-differentiated species, regardless of the data set and analyses employed. Our CHC and cox1-based network showed the members of E. ruidum sp. 1 as a highly divergent cluster whose geographic distribution ranges from southeast Mexico to Ecuador, Colombia and Venezuela in South America, and in the Lesser Antilles in the Caribbean. This network also showed the existence of geographic structure within this species. The name E. ruidum should be applied to the populations assigned to E. ruidum sp. 1 based on the type locality of the species, which was restricted to Colombia by Kugler and Brown (1982), and on their morphological correspondence with the syntypes from this country (Aguilar-Velasco et al. 2016). The 3RAD, UCE and cox1 analyses also recovered the specimens of E. ruidum sp. 2 as a separate species, being composed of two geographically structured clades. One of these clades included specimens from southeast Mexico to Colombia and Ecuador in South America, whereas the other one was represented by specimens from southeast to northern Mexico. The cox1 data set, however, showed a considerable genetic distance between these clusters, suggesting that they could be separate species. This hypothesis is supported by the CHC analyses, which also showed a marked divergence between the two clades.

The remaining three species delimited here correspond to populations from the lowlands of Sierra Madre del Sur in the states of Oaxaca and Guerrero, in southeast Mexico. The two genome-wide data sets and the CHC profiles indicated that *E. ruidum* sp. 3 and *E. ruidum* sp. 4 are two closely related species whose geographic distribution is restricted to lowland areas of Oaxaca. Moreover, our results consistently support the hypothesis that the specimens of the putative hybrid population proposed by Aguilar-Velasco et al. (2016) from Pinotepa Nacional, Oaxaca, and those assigned to *E. ruidum* sp. 3 from the state of Guerrero, actually represent a distinct species. Further comprehensive sampling on foothills along the Pacific coast in southeast Mexico will reveal the actual geographic distribution of these three species. Moreover, taxonomic inferences based on these results should consider the previous name availability within the group. *Ectatomma aztecum* was described by Emery (1901) based on a single specimen collected in the state of Michoacán, Mexico, but without precise locality; however, it was subsequently regarded as a synonym of *E. ruidum* by Kugler and Brown (1982). According to Aguilar-Velasco et al. (2016) the syntype of *E. aztecum*(CASENT0903841; MSNG, Genoa, Italy) is morphologically similar to the specimens assigned to *E. ruidum* sp. 3.

Potential drivers of speciation in the E. ruidum complex

Allopatric speciation, which involves geographic isolation of segments of a formerly contiguous population, is firmly established as the primary mechanism by which new species evolve (Coyne and Orr 2004). Without spatial isolation among populations, it is difficult to have a scenario in which gene flow can be reduced or eliminated. However, it is known that speciation can be multifactorial with multiple axes of differentiation characterizing even incipient species (Mullen and Shaw 2014; Brodetzki et al. 2019). For instance, in *Heliconius* butterflies color pattern divergence, mate preference, host-plant use and microhabitat choice possibly are implicated in species diversification (Smiley 1978; Brown Jr 1981; Estrada and Jiggins 2002; Mallet et al. 2007). In the case of the species of the *E. ruidum* complex, we do not see any clear correlation between their distribution pattern and geographic barriers, nor possible events of microallopatry or niche specialization. For instance, in the state of Chiapas, Mexico, the populations of *E. ruidum* sp. 1 and sp. 2 occurred in quasi-sympatry, at less than 15 km from each other (Lachaud 1990). Even more puzzling is the geographic distribution of the three delimited species restricted to lowland areas of Oaxaca and Guerrero, in southeast Mexico, which we found to be separated from each other only by 5 to 30 km. Below we discuss three potential phenotypic and genotypic factors that could have promoted the speciation among the species of the *E. ruidum* complex: mitonuclear conflict, chemical cue divergence, and colony structure.

Mt performance affects every aspect of individual fitness and its correct function relies on mitonuclear compatibility (Angers et al. 2018; Zaidi and Makova 2019). This intimate interaction between the mt and nuclear products has led to a strong coevolution between both genomes (Blier et al. 2001; Burton et al. 2013; Chou and Leu 2015). It has been proposed that mitonuclear coevolution in isolated populations triggers speciation, since population-specific mitonuclear coadaptations lead to between-population mitonuclear incompatibility, thus precluding gene flow (Hill 2016). It is therefore plausible that the rapidly evolving mt haplotypes found in the three heteroplasmic species of the *E. ruidum* complex restricted to southeast Mexico played a role in their speciation if they were expressed and interacted with the nuclear products, consequently pressing for its evolution (Hill 2017). This could have been achieved through a marked population structure promoted by assortative mating, which created mitonuclear incompatibilities and barriers to gene flow between the taxa involved. However, since heteroplasmy does not seem to occur in the widely distributed *E.ruidum* spp. 1 and 2, other factors probably could also have promoted the species diversification in this complex.

Social behaviour and recognition cues could act as major traits that could drive assortative mating, limit gene flow and ultimately promote speciation (Hochberg et al. 2003). Sympatric speciation based on social trait divergence has been suggested for ants of the *Cataglyphis niger* complex, where three incipient species possess consistent differences in CHC composition, social structure and mtDNA sequence data (Brodetzki et al. 2019). Our results showed that the CHC distances among populations mirrored their phylogenetic relationships obtained from mt and genome-wide data. A similarly high correlation between CHC and genetic data has been observed in other social insects (Isoptera: Dronnet et al. 2006; stingless bees: Leonhardt et al. 2013; ants: Hartke et al. 2019). CHC divergence has been associated with nestmate/non-nestmate communication (van Zweden and d'Etorre, 2010) and may also play an important role in pre-mating isolation (Savarit et al. 1999; Smadja and Butlin 2009; Snellings et al. 2018).

Another important trait in social insects is colony structure. In some ant groups, shifts from monogynous to polygynous colonies and gyne morphological polymorphism (macro and microgynes) are also known to be involved in their speciation (Seifert 2010; Brodetzki et al. 2019). For instance, limited dispersal of females of polygynous species can promote differentiation between populations, and if male dispersal is also restricted this can also lead to their speciation (Pamilo and Rosengren 1984; Seppä and Pamilo 1995; Gyllenstrand et al. 2002). A previous study that focused on a population of *E. ruidum* sp. 2 from Rosario Izapa in the state of Chiapas, southeast Mexico, showed that macrogynes and microgynes can be adopted by their monogynous and polygynous mother colonies, leading to low dispersal (Lenoir et al. 2011). Moreover, queen size dimorphism and social polymorphism has been observed in laboratory colonies both in *E. ruidum* sp. 3 and sp. 4 (Peña-Carrillo, unpublished data), where few microgynes were produced. Further colony structure studies performed for species of the *E. ruidum* complex will reveal whether the presence of queen size dimorphism and social polymorphism was implicated in their diversification process.

Ectatomma ruidum has been the subject of a vast number of studies, including those on social structure (e.g. Corbara et al. 1989), foraging and diet (e.g. Lachaud 1990; Riera-Valera and Pérez-Sánchez 2009; Santamaría et al. 2009), macronutrient regulation (e.g. Cook and Behmer 2010), home ranges and nestmate recognition (e.g. Breed et al. 1990), queen dimorphism (e.g. Lachaud et al. 1999; Lenoir et al. 2011), communication behavior (e.g. Pratt 1989), and parasitoid interaction (Howard et al. 2001). Most of these studies, however, were based on specimens from different localities across the Neotropics. Our gathered evidence consistently

shows that *E. ruidum* actually represents a species-complex of species, which would change the interpretation and the extension of the conclusions that were drawn in previous studies. Our study thus highlights the importance of using different sources of molecular data for species delimitation of morphologically conserved taxa. Moreover, it is important to consider the occurrence of heteroplasmy in systematic studies, since it can lead to incorrect estimates of phylogenies if it is not detected. The presence of extensive heteroplasmy within the *E. ruidum* complex highlights the necessity of being aware of the occurrence of this phenomenon in other insect groups and metazoans in general, since it appears that it is not as rare as previously thought (White et al 2008; Robison et al. 2015; Macey et al. 2021).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments.

We thank Michael Branstetter and Natalia Bayona for their help in the laboratory to generate the UCE and 3RAD data, respectively; Jean-Paul Lachaud, Carlos Santamaria and Gabriela Pérez-Lachaud for their help during the field trips to Mexico in the state of Oaxaca and Colombia. We also thank Cristina Mayorga and Guillermina Ortega for their help with the curation of specimens at the CNIN IBUNAM; Laura Márquez, Nelly López and Andrea Jiménez for their assistance in the laboratory; Ian Butler for his comments to the text; and the Direccion General de Computo y de Tecnologias de Informacion y Comunicacion (DGTIC, UNAM) for the services and resources provided to run part of the analyses in the Miztli supercomputer (Proyecto de investigacion regular LANCAD UNAM-DGTIC-339). This work was funded by grants given by DGAPA UNAM (Proyecto PAPIIT convocatoria 2019, project number IN201119) and CONACyT (Proyecto Ciencia de Frontera 2019 No. 58548) to AZR. RNML was supported by a postdoctoral grant given by the Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México, and KPC by a PhD scholarship given by CONACyT and the French government.

Data availability statement

3RAD and UCE raw reads generated in this study are deposited at the National Center for Biotechnology Sequence Read Archive (SRA; Bioproject ID XXXX). All the matrices that were analyzed in the study and their resulting topologies are available in the Fogshare repository XXXXXX.

References

Aberer, A. J., Kobert, K., and Stamatakis, A. (2014). Exabayes: Massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution* **31**, 2553–2556. doi:10.1093/molbev/msu236

Adams, S. A., and Tsutsui, N. D. (2020). The evolution of species recognition labels in insects. *Philosophical Transactions of the Royal Society B* **375**, 20190476. doi:10.1098/rstb.2019.0476

Aguilar-Velasco, R. G., Poteaux, C., Meza-Lázaro, R., Lachaud, J-P., Dubovikoff, D., and Zaldívar-Riverón, A. (2016). Uncovering species boundaries in the Neotropical ant complex *Ectatomma ruidum*(Ectatomminae) under the presence of nuclear mitochondrial paralogues. *Zoological Journal of the Linnean Society* **178**, 226–240. doi:10.1111/zoj.12407

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B. C., and Antonelli, A. (2019). Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements. *Systematic Biology* **68**, 32–46. doi:10.1093/sysbio/syy039

Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online, 1–15. doi:10.1002/9781118445112.stat07841

Angers, B., Leung, C., Vétil, R., Deremiens, L., and Vergilino, R. (2018). The effects of allospecific mitochondrial genome on the fitness of northern redbelly dace (Chrosomus eos). *Ecology and Evolution***8**, 3311–3321. doi:10.1002/ece3.3922 Antweb. Version 8.64.2. California Academy of Science, online at https://www.antweb.org. Accessed 4 September 2021.

Bagnères, A.G. and Wicker-Thomas C (2010) Chemical taxonomy with hydrocarbons. In: Blomquist, G.J. and Bagnères, A.G. (eds). Insect hydrocarbons: biology, biochemistry and chemical ecology. Cambridge University Press, Cambridge, pp 121–162.

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., and Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* **2019**, 1–25. doi:10.7717/peerj.7724

Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences* **92**, 11331-11338. doi:10.1073/pnas.92.25.11331

Blier, P. U., Dufresne, F., and Burton, R. S. (2001). Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *TRENDS in Genetics* **17**, 400–406. doi:10.1016/s0168-9525(01)02338-1

Blomquist, G. J., and Bagnères, A.G. (2010). 'Insect hydrocarbons: biology, biochemistry, and chemical ecology'. (Cambridge University Press.)

Boomsma, J. J., and Nash, D. R. (2014). Evolution: sympatric speciation the eusocial way. *Current Biology* 24, R798–R800. doi:10.1016/j.cub.2014.07.072

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* **10**, 1–6. doi:10.1371/journal.pcbi.1003537

Bradbury, J. W., Vehrencamp, S. L., and others (1998). Principles of animal communication. Sinauer Associates Inc.

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., and Brady, S. G. (2017). Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. *Current Biology* **27**, 1019–1025. doi:10.1016/j.cub.2017.03.027

Breed, M. D., Abel, P., Bleuze, T. J., and Denton, S. E. (1990). Thievery, home ranges, and nestmate recognition in *Ectatomma ruidum*. *Oecologia* 84, 117-121. doi:10.1007/BF00665604

Brodetzki, T. R., Inbar, S., Cohen, P., Aron, S., Privman, E., and Hefetz, A. (2019). the Interplay between Incipient species and social polymorphism in the Desert Ant Cataglyphis. *Scientific Reports***9**, 1–14.

Brown Jr, K. S. (1981). The biology of Heliconius and related genera. Annual Review of Entomology 26, 427–457.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., and Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**, 1917–1932. doi:10.1093/molbev/mss086

Bryant, D., and Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255-265. doi:10.1093/molbev/msh018

Burton, R. S., and Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology***21**, 4942–4957. doi:10.1111/mec.12006

Burton, R. S., Pereira, R. J., and Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. Annual Review of Ecology, Evolution, and Systematics 44, 281–302. doi:10.1146/annurev-ecolsys-110512-135758 Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and Evolution* **17**, 540–552. doi:10.1093/oxfordjournals.molbev.a026334

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**, 3124–3140. doi:10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes General Genetics 1, 171–182. doi:10.1534/g3.111.000240

Chifman, J., and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioin-formatics* **30**, 3317–3324. doi:10.1093/bioinformatics/btu530

Chou, J. Y., and Leu, J. Y. (2010). Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *BioEssays* **32**, 401–411. doi:10.1002/bies.200900162

Chou, J. Y., and Leu, J. Y. (2015). The Red Queen in mitochondria: Cyto-nuclear co-evolution, hybrid breakdown and human disease. *Frontiers in Genetics* **6**, 1–8. doi:10.3389/fgene.2015.00187

Chung, H., and Carroll, S. B. (2015). Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. *BioEssays* **37**, 822–830. doi:10.1002/bies.201500014

Clement, M., Snell, Q., Walker, P., Posada, D., and Crandall, K. (1997). TCS networks.pdf.

Cook, S. C., and Behmer, S. T. (2010). Macronutrient regulation in the tropical terrestrial ant Ectatomma ruidum (Formicidae): A field study in Costa Rica. *Biotropica* 42, 135–139. doi:10.1111/j.1744-7429.2009.00616.x

Corbara, B., Lachaud, J-P., and Fresneau, D. (1989). Individual Variability, Social Structure and Division of Labour in the Ponerine Ant Ectatomma ruidum Roger (Hymenoptera, Formicidae). *Ethology*82, 89–100. doi:10.1111/j.1439-0310.1989.tb00490.x

Coyne, J. A. and Orr, H. A. (2004). 'Speciation'. (Sinauer Associates Sunderland, MA.)

Crespi, B., and Nosil, P. (2013). Conflictual speciation: Species formation via genomic conflict. Trends in Ecology and Evolution 28, 48–57. doi:10.1016/j.tree.2012.08.015

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85**, 407–417. doi: 10.1111/j.1095-8312.2005.00503.x

Dronnet, S., Lohou, C., Christides, J.-P., and Bagneres, A.-G. (2006). Cuticular hydrocarbon composition reflects genetic relationship among colonies of the introduced termite *Reticulitermes santonensis* Feytaud. *Journal of Chemical Ecology* **32**, 1027–1042.

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844–1849. doi:10.1093/bioinformatics/btu121

Eaton, D. A. R., and Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594. doi:10.1093/bioinformatics/btz966

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797. doi:10.1093/nar/gkh340

Edwards, D. L., and Knowles, L. L. (2014). Species detection and individual assignment in species delimitation: Can integrative data increase efficacy? *Proceedings of the Royal Society B: Biological Sciences* **281**. doi:10.1098/rspb.2013.2765

Emery, C. (1901). Notes sur les sous-familles des Dorylines et Ponerines (Famille des Formicides). Annales de la Societe entomologique de Belgique 45: 32-54

Estrada, C., and Jiggins, C. D. (2002). Patterns of pollen feeding and habitat preference among Heliconius species. *Ecological Entomology* **27**, 448–456. doi:10.1046/j.1365-2311.2002.00434.x

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620.

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods in Ecology and Evolution 8, 1103–1112. doi:10.1111/2041-210X.12754

Faircloth, B. C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. See http://dx. doi. org/10.6079/J9ILL (accessed 24 September 2021).

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioin-formatics* **32**, 786–788. doi:10.1093/bioinformatics/btv646

Faircloth, B. C. (2021). Tutorial II: Phasing UCE data. https://phyluce.readthedocs.io/en/latest/tutorials/tutorial-2.html.

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., and Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* **61**, 717–726. doi:10.1093/sysbio/sys004

Flouri, T., Jiao, X., Rannala, B., and Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution* **35**, 2585–2593. doi:10.1093/molbev/msy147

Gershoni, M., Templeton, A. R., and Mishmar, D. (2009). Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays***31**, 642–650. doi: 10.1002/bies.200800139

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vasquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., and others (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru \& iNext). *PeerJ* 7, e7755. doi:10.7717/peerj.7755

Grummer, J. A., Bryson, R. W., and Reeder, T. W. (2014). Species delimitation using bayes factors: Simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* **63**, 119–133. doi:10.1093/sysbio/syt069

Grun, P. (1976). 'Cytoplasmic genetics and evolution'. (Columbia University Press.)

Gyllenstrand, N., Gertsch, P. J., and Pamilo, P. (2002). Polymorphic microsatellite DNA markers in the ant *Formica exsecta .Molecular Ecology Notes* **2**, 67–69. doi:10.1046/j.1471-8286.2002.00152.x

Hakala, S. M., Perttu, S., and Helantera, H. (2019). Evolution of dispersal in ants (Hymenoptera: Formicidae): A review on the dispersal strategies of sessile superorganisms. $Myrmecological \ news 29$. doi:10.25849/myrmecol.news_029:035

Harrison, R. G., and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity***105**, 795–809. doi:10.1093/jhered/esu033

Hartke, J., Sprenger, P. P., Sahm, J., Winterberg, H., Orivel, J., Baur, H., Beuerle, T., Schmitt, T., Feldmeyer, B., and Menzel, F. (2019). Cuticular hydrocarbons as potential mediators of cryptic species divergence in a mutualistic ant association. *Ecology and Evolution* **9**, 9160–9176. doi:10.1002/ece3.5464

Hebert, P. D. N., Cywinska, A., Ball, S. L., and Dewaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences***270**, 313–321. doi:10.1098/rspb.2002.2218

Hill, G. E. (2016). Mitonuclear coevolution as the genesis of speciation and the mitochondrial DNA barcode gap. *Ecology and Evolution* $\mathbf{6}$, 5831–5842. doi:10.1002/ece3.2338

Hill, G. E. (2015). Mitonuclear ecology. *Molecular Biology and Evolution* **32**, 1917–1927. doi:10.1093/molbev/msv104

Hill, G. E. (2018). Mitonuclear Mate Choice: A Missing Component of Sexual Selection Theory? *BioEssays* **40**, 1–10. doi:10.1002/bies.201700191

Hill, G. E. (2019). Reconciling the mitonuclear compatibility species concept with rampant mitochondrial introgression. *Integrative and Comparative Biology* **59**, 912–924. doi:10.1093/icb/icz019

Hill, G. E. (2017). The mitonuclear compatibility species concept. *The Auk: Ornithological Advances* **134**, 393–409. doi:10.1642/AUK-16-201.1

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UF-Boot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518-522. doi:10.1093/molbev/msx281

Hochberg, M. E., Sinervo, B., and Brown, S. P. (2003). Socially mediated speciation. Evolution, 57, 154-158. doi:10.1111/j.0014-3820.2003.tb00224.x

Howard, R. W., Perez-Lachaud, G. and Lachaud, J-P. (2001) Cuticular hydrocarbons of *Kapala sulci-facies* (Hymenoptera: Eucharitidae) and its host, the ponerine ant *Ectatomma ruidum* (Hymenoptera: Formicidae). *Annals of the Entomological Society of America***94** : 707-716. doi:10.1603/0013-8746(2001)094[0707:CHOKSH]2.0.CO;2

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution***23**, 254-267. doi:10.1093/molbev/msj030

Ilut, D. C., Nydam, M. L., and Hare, M. P. (2014). Defining loci in restriction-based reduced representation genomic data from nonmodel species: sources of bias and diagnostics for optimal clustering. *BioMed Research International*, 2014. doi:10.1155/2014/675158

Johnstone, R. A., Cant, M. A., and Field, J. (2012). Sex-biased dispersal, haplodiploidy and the evolution of helping in social insects. *Proceedings of the Royal Society B: Biological Sciences***279**, 787–793. doi:10.1098/rspb.2011.1257

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**, 1160–1166. doi:10.1093/bib/bbx108

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589. doi: 10.1038/nmeth.4285

Kugler, C., and Brown, W. L. (1982). Revisionary and other studies on the ant genus Ectatomma, including the descriptions of two new species. Estudios de revision y de otros tipos en el genero de hormigas Ectatomma, incluyendo las descripciones de dos nuevas especies. Agriculture 24, 1–7.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* **35**, 1547–1549. doi:10.1093/molbev/msy096

Lachaud, J.-P. (1990). Foraging activity and diet in some neotropical ponerine ants. I. *Ectatomma ruidum* Roger (Hymenoptera, Formicidae). *Foliae Entomologica Mexicana* **78**, 241–256.

Lachaud, J. P., Cadena, A., Schatz, B., Perez-Lachaud, G., and Ibarra-Nunez, G. (1999). Queen dimorphism and reproductive capacity in the ponerine ant, Ectatomma ruidum Roger. *Oecologia* **120**, 515–523. doi:10.1007/s004420050885

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772–773. doi:10.1093/molbev/msw260

Leache, A. D., and Bouckaert, R. R. (2018). Species trees and species delimitation with SNAPP: a tutorial and worked example. In Workshop on Population and Speciation Genomics, Český Krumlov.

Leaché, A. D., and Fujita, M. K. (2010). Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society B: Biological Sciences***277**, 3071-3077. doi:10.1098/rspb.2010.0662

Leaché, A. D., Fujita, M. K., Minin, V. N., and Bouckaert, R. R. (2014). Species delimitation using genomewide SNP Data. *Systematic Biology* **63**, 534–542. doi:10.1093/sysbio/syu018

Leavitt, S. D., Moreau, C. S., and Lumbsch, H. T. (2015). The dynamic discipline of species delimitation: progress toward effectively recognizing species boundaries in natural populations. In 'Recent advances in lichenology'. pp. 11–44. (Springer.)

Leigh, J. W., and Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. Methods in Ecology and Evolution6, 1110–1116. doi:10.1111/2041-210X.12410

Lenoir, J.C., Lachaud, J.P., Nettel, A., Fresneau, D., Poteaux, C. (2011). The role of the microgynes in the reproductive strategy of the neotropical ant *Ectatomma ruidum*. *Naturwissenschaften*. 98:347–356. doi:10.1007/s00114-011-0774-3

Leonhardt, S. D., Rasmussen, C., and Schmitt, T. (2013). Genes versus environment: Geography and phylogenetic relationships shape the chemical profiles of stingless bees on a global scale. *Proceedings of the Royal Society B: Biological Sciences* **280**, 7–9. doi:10.1098/rspb.2013.0680

Lucas, C., Fresneau, D., Kolmer, K., Heinze, J., Delabie, J. H. C., and Pho, D. B. (2002). A multidisciplinary approach to discriminating different taxa in the species complex *Pachycondyla villosa*(Formicidae). *Biological Journal of the Linnean Society***75**, 249–259. doi:10.1046/j.1095-8312.2002.00017.x

Macey, J. R., Pabinger, S., Barbieri, C. G., Buring, E. S., Gonzalez, V. L., Mulcahy, D. G., DeMeo, D. P., Urban, L., Hime, P. M., Prost, S., Elliot, A. N., and Gemmel, N. J. (2021). Evidence of two deeply divergent co-existing mitochondrial genomes in the Tuatara reveals an extremely complex genomic organization. *Communications Biology* 4, 116. doi:10.1038/s42003-020-01639-0

Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. (2007). Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology* **7**, 1–16. doi:10.1186/1471-2148-7-28

Meza-Lázaro, R. N., Poteaux, C., Bayona-Vásquez, N. J., Branstetter, M. G., and Zaldívar-Riverón, A. (2018). Extensive mitochondrial heteroplasmy in the neotropical ants of the *Ectatomma ruidum* complex (Formicidae: Ectatomminae). *Mitochondrial DNA Part A*29, 1203–1214. doi:10.1080/24701394.2018.1431228

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop, GCE 2010. doi:10.1109/GCE.2010.5676129

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., S. Swenson, M., and Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, 541–548. doi:10.1093/bioinformatics/btu462

Mirarab, S., and Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, i44–i52. doi:10.1093/bioinformatics/btv234

Mullen, S. P., and Shaw, K. L. (2014). Insect speciation rules: unifying concepts in speciation research. Annual Review of Entomology 59, 339–361. doi:10.1146/annurev-ento-120710-100621

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32(1), 268-274. doi:10.1093/molbev/msu300

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2015). Vegan: Community Ecology Package. R package version 2.0-10. 2013.

Padial, J. M., Miralles, A., la Riva, I., and Vences, M. (2010). The integrative future of taxonomy. *Frontiers in zoology* **7**, 1–14. doi: 10.1186/1742-9994-7-16

Pamilo, P., and Rosengren, R. (1984). Evolution of nesting strategies of ants: genetic evidence from different population types of *Formica*ants. *Biological Journal of the Linnean Society* **21**, 331–348. doi: 10.1111/j.1095-8312.1984.tb00370.x

Pante, E., Schoelinck, C., and Puillandre, N. (2015). From integrative taxonomy to species description: one step beyond. *Systematic Biology* **64**, 152–160. doi: 10.1093/sysbio/syu083

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., and Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology* **17**, 337–354. doi:10.1089/cmb.2009.0179

Peña-Carrillo, K. I., Poteaux, C., Leroy, C., Meza-Lázaro, R. N., Lachaud, J. P., Zaldívar-Riverón, A., and Lorenzi, M. C. (2021a). Highly divergent cuticular hydrocarbon profiles in the clepotobiotic ants of the *Ectatomma ruidum* species complex. *Chemoecology***31**, 125–135. doi:10.1007/s00049-020-00334-0.

Peña-Carrillo, K. I., Lorenzi, M. C., Brault, M., Devienne, P., Lachaud, J. P., Pavan, G., Poteaux, C. (2021b). A new putative species in the *Ectatomma ruidum* complex (Formicidae: Ectatomminae) produces a species-specific distress call. *Bioacustics*, DOI: 10.1080/09524622.2021.1938226.

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., and Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology and Evolution* **25**, 459–467. doi:10.1016/j.tree.2010.05.006

Pratt, S. C. (1989). Recruitment and other communication behavior in the ponerine ant *Ectatomma ruidum* . *Ethology* **81**, 313–331. doi: 10.1111/j.1439-0310.1989.tb00777.x

Presgraves, D. C. (2010). Speciation genetics: Search for the missing snowball. *Current Biology* **20**, R1073–R1074. doi:10.1016/j.cub.2010.10.056

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*155, 945–959. doi:10.1093/genetics/155.2.945

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic biology ${\bf 56}$, 879–886. doi:10.1080/10635150701701083

Radzvilavicius, A. L., Lane, N., and Pomiankowski, A. (2017). Sexual conflict explains the extraordinary diversity of mechanisms regulating mitochondrial inheritance. *BMC Biology* **15**, 1–12. doi:10.1186/s12915-017-0437-8

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization on Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901-904. doi: 10.1093/sysbio/syy032.

Riera-Valera, M. A., and Pérez-Sánchez, A. J. (2009). Notas acerca de la dieta de *Ectatomma ruidum* (Roger 1861) (Hymenoptera: Formicidae: Ectatomminae) en un jardín venezolano. *Boletín Sociedad Entomológica Aragonesa* 44, 550–552.

Robison, G. A., Balvin, O., Schal, C., Vargo, E. L., and Booth, W. (2015). Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). *Journal of* Medical Entomology 52, 734-738. doi: 10.1093/jme/tjv055

Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome research* **22**, 939-946. doi:10.1101/gr.128124.111

Roger, J. (1860). Die Ponera-artigen Ameisen. Berliner Entomologische Zeitschrift 4, 278–312.

Santamaría, C., Domínguez Haydar, Y., and Armbrecht, I. (2009). Cambios en la distribución de nidos y abundancia de la hormiga *Ectatomma ruidum* (Roger 1861) en dos zonas de Colombia. *Boletín del Museo de Entomología de la Universidad del Valle* **10**, 10–18.

Sato, M., and Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta - Molecular Cell Research* **1833**, 1979–1984. doi:10.1016/j.bbamcr.2013.03.010

Savarit, F., Sureau, G., Cobb, M., and Ferveur, J. F. (1999). Genetic elimination of known pheromones reveals the fundamental chemical bases of mating and isolation in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9015–9020. doi:10.1073/pnas.96.16.9015

Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R. H. (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual review of entomology* **55**, 421–438. doi: 10.1146/annurev-ento-112408-085432

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., et al. (2014). Genomics and the origin of species. *Nature Reviews Genetics* **15**, 176–192. doi:10.1038/nrg3644

Seifert, B. (2010). Intranidal mating, gyne polymorphism, polygyny, and supercoloniality as factors for sympatric and parapatric speciation in ants. *Ecological Entomology* **35**, 33–40.

Seo, T. K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution***25**, 960–971. doi:10.1093/molbev/msn043

Seppä, P., and Pamilo, P. (1995). Gene flow and population viscosity in Myrmica ants. *Heredity*, 74(2), 200-209. doi: 10.1038/hdy.1995.28

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome research* **19**, 1117–1123. doi:10.1101/gr.089532.108

Sloan, D. B. (2015). Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist* **205**, 1040–1046. doi:10.1111/nph.12835

Smadja, C., and Butlin, R. K. (2009). On the scent of speciation: The chemosensory system and its role in premating isolation. *Heredity***102**, 77–97. doi:10.1038/hdy.2008.55

Smiley, J. (1978). Plant chemistry and the evolution of host specificity: new evidence from Heliconius and Passiflora. *Science*201, 745–747. doi:10.1126/science.201.4357.745

Snellings, Y., Herrera, B., Wildemann, B., Beelen, M., Zwarts, L., Wenseleers, T., and Callaerts, P. (2018). The role of cuticular hydrocarbons in mate recognition in Drosophila suzukii. *Scientific Reports* **8**, 1–11. doi:10.1038/s41598-018-23189-6

Song, H., Moulton, M. J., and Whiting, M. F. (2014). Rampant nuclear insertion of mtDNA across diverse lineages with in Orthoptera (Insecta). *PLoS ONE* **9**, 41–43. doi:10.1371/journal.pone.0110508

Sprenger, P. P., and Menzel, F. 2020. Cuticular hydrocarbons in ants (Hymenoptera: Formicidae) and other insects: how and why they differ among individuals, colonies, and species. Myrmecol. News 30: 1-26. doi: 10.25849/myrmecol.news_030:001

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogeneis. *Bioinformatics* **30**, 1312–1313. doi:10.1093/bioinformatics/btu033

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic biology 56, 564-577. doi:10.1080/10635150701472164

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., and Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic acids research*, 44(W1), W232-W235. doi:10.1093/nar/gkw256

Will, K. W., Mishler, B. D., and Wheeler, Q. D. (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Systematic biology* **54**, 844–851. doi:10.1080/10635150500354878

Wollenberg Valero, K. C., Marshall, J. C., Bastiaans, E., Caccone, A., Camargo, A., Morando, M., Niemiller, M. L., Pabijan, M., Russello, M. A., Sinervo, B., and others (2019). Patterns, mechanisms and genetics of speciation in reptiles and amphibians. *Genes* **10**, 646. doi:10.3390/genes10090646

Yang, Z., and Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9264–9269. doi:10.1073/pnas.0913022107

Yang, Z., and Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution* **31**, 3125–3135. doi:10.1093/molbev/msu279

Zaidi, A. A., and Makova, K. D. (2019). Investigating mitonuclear interactions in human admixed populations. *Nature ecology* \and evolution **3**, 213–222. 10.1038/s41559-018-0766-1

van Zweden JS, d'Ettorre P (2010) Nestmate recognition in social insects and the role of hydrocarbons. In: Blomquist GJ, Bagnères A-G (eds) Insect hydrocarbons: biology, biochemistry and chemical ecology. Cambridge University Press, Cambridge, pp 222–243.

White, D. J., Wolff, J. N., Pierson, M., and Gemmell, N. J. (2008). Revealing the hidden complexities of mtDNA inheritance. Molecular Ecology, 17(23), 4925-4942. doi: 10.1111/j.1365-294X.2008.03982.x

Figure legends

Figure 1. Maps showing the sampled localities for the putative species belonging to the *E. ruidum* complex that were delimited in this study: a) red: *E. ruidum* sp. 1; b) green: *E. ruidum*sp. 2; c) yellow: *E. ruidum* sp. 3; violet: *E. ruidum* sp. 4; black: *E. ruidum* sp. "Pinotepa + Guerrero".

Figure 2. Summary of the results obtained with the different sources of information examined in this study: a) hierarchical cluster dendrogram built with the dataset of the simplified CHCs profile of the *E. ruidum* species complex; b) bayesian phylogram derived from the 3RAD dataset (minimum sample locus = 30); c) ML phylogram derived from the UCEs dataset (95% taxon occupancy matrix); d) haplotype network reconstructed with the *cox1* dataset.

Figure 3. Patterns of genetic structure and admixture among the examined populations based on the 3RAD dataset (98_33 matrix) for values of population differentiation (K) from 2 to 8.

Figure 4. Pearson's product-moment correlation between CHC and 3RAD, UCEs and cox1 distances.

Supplementary Material

Table S1. List of examined specimens, their taxon ID, locality, GenBank accession number and number of colonies for the 3RAD, UCE, *cox1* and CHC datasets.

Table S2. Statistics for different percentages of minimum taxon coverage for the examined 3RAD datasets. PIS= parsimony informative sites.

Table S3. Statistics for different percentages of minimum taxon coverage for the examined UCE datasets. PIS= parsimony informative sites.

Table S4. BPP Species delimitation based on 200 (A) and 50 (B) loci matrices with three and ten replicates (r), respectively. Each number (1 or 0) represents one of the nodes, from 10 to 17, according to the guide tree. Zeros represent collapsed nodes. Red strings have at least one collapsed node. PP= posterior probabilities [?] 0.1. The columns represent three sets of parameters varying the ancestral population size (ϑ) and root age (τ).

Figure S1. ML phylogram derived from the *cox1* dataset with ultrafast bootstrap values (UFBoot2).

Figure S2. Bayesian and ML phylograms derived from the 3RAD and UCEs datasets, and the unrooted phylogenetic network derived from the alignment of the concatenated UCE phased loci.







| ar | |
|--|--|
| | |
| | |
| hid | |
| pe | |
| ay | |
| В | |
| ata | |
| Ĥ | |
| ed. | |
| ew | |
| evi | |
| 10 | |
| be | |
| en | |
| p, | |
| lot | |
| 38 | |
| Ĩ. | |
| anc | |
| f | |
| | |
| ore] | |
| a I | |
| nis | |
| E | |
| | |
| 7 | |
| 13 | |
| 454 | |
| 026 | |
| 1.46 | |
| 33.1 | |
| 121 | |
| 334 | |
| Ē | |
| a | |
| 541 | |
| 225 | |
| | |
| 60 | |
| 1.01 | |
| qo | |
| 1 | |
| tps | |
| Ę | |
| | |
| | |
| on. – | |
| ssion | |
| mission | |
| permission. — | |
| ut permission. — | |
| chout permission. — | |
| without permission. — | |
| se without permission. — | |
| reuse without permission. — | |
| Vo reuse without permission. — | |
| 1. No reuse without permission. — | |
| ved. No reuse without permission. — | |
| served. No reuse without permission. — | |
| reserved. No reuse without permission. — | |
| chts reserved. No reuse without permission. — | |
| rights reserved. No reuse without permission. — | |
| All rights reserved. No reuse without permission. — | |
| r. All rights reserved. No reuse without permission. — | |
| nder. All rights reserved. No reuse without permission. — | |
| funder. All rights reserved. No reuse without permission. — | |
| or/funder. All rights reserved. No reuse without permission. — | |
| uthor/funder. All rights reserved. No reuse without permission. — | |
| author/funder. All rights reserved. No reuse without permission. — | |
| the author/funder. All rights reserved. No reuse without permission. — | |
| is the author/funder. All rights reserved. No reuse without permission. — | |
| der is the author/funder. All rights reserved. No reuse without permission. — | |
| nolder is the author/funder. All rights reserved. No reuse without permission. — | |
| it holder is the author/funder. All rights reserved. No reuse without permission. — | |
| ight holder is the author/funder. All rights reserved. No reuse without permission. — | |
| pyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| - The copyright holder is the author/funder. All rights reserved. No reuse without permission. | |
| 1- The copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| 021 - The copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| t 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| 0ct 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| 14 Oct 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| ea 14 Oct $2021 - $ The copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| norea 14 Oct $2021 - $ The copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| uthorea 14 Oct $2021 - $ The copyright holder is the author/funder. All rights reserved. No reuse without permission. $-$ | |
| ı Authorea 14 Oct 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| on Authorea 14 Oct 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |
| ted on Authorea 14 Oct 2021 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — | |

