Molecular Phylogenetics and Evolution 115 (2017) 128–139

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

An integrative approach to untangling species delimitation in the *Cataglyphis bicolor* desert ant complex in Israel

P.A. Eyer*, R. Seltzer, T. Reiner-Brodetzki, A. Hefetz

Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Keywords:
Species delimitation
Phylogeography
Social structure
Ants
Cataglyphis

ABSTRACT

Although extensive research has been carried out on the desert ants in the genus *Cataglyphis* in recent years, some of the specific intra- and interspecific relationships remain elusive. The present study disentangles the phylogenetic relationships among the *C. bicolor* complex in Israel using an integrative approach based on genetic markers, morphometric measurements, and chemical analyses (cuticular hydrocarbons). Several species delimitation approaches based on four nuclear, two mitochondrial, and eleven microsatellite markers, as well as 16 body measurements and 56 chemical variables, were employed to deciphering the occurrence of cryptic species in our data set.

Our findings support the occurrence of at least four distinct species in the *C. bicolor* group in Israel, one of which may be a complex of three more recent species. The findings confirm the distinctiveness of *C. isis* and *C. holgerseni*. They attest the presence of a recently discovered species, *C. israelensis,* in the central mountain ridge and the occurrence of another clade distributed from the Negev to the Mediterranean coast, comprising the species *C. niger, C. savignyi,* and *C. drusus*. Although these three species are separated on the basis of *mt*DNA, this subgrouping was not supported by any of the nuclear sequence markers nor by the microsatellite analysis. This genetic structure may thus either reflect a possible recent speciation, or a geographical structuring of a single species. Overall, using these different sources of evidence we locate our samples within a global phylogeny of the *bicolor* group and discuss the processes that underlie speciation in this group.

1. Introduction

Species constitute fundamental units in evolutionary biology, with their delimitation, therefore, having profound implications for the interpretation of biological phenomena, from physiology to community ecology and evolution (Wiens, 2007; Schlick-Steiner et al., 2010). The more traditional mode of classification based on phenotype criteria has come into question in view of the increasing number of studies reporting the occurrence of morphologically cryptic species (Bickford et al., 2007). In those cases molecular methods were used to circumvent the difficulties of morphology-based delimitation. To date, it is widely accepted that (*i*) species form separately evolving lineages, the members of which share tighter genealogical relations to each other than to members of different species (de Queiroz, 2007); and (*ii*) populations of different species are able to coexist in sympatry while maintaining their distinctiveness through reproductive isolation (Mayr, 1942). From this perspective, molecular species delimitation using genetic markers aims at identifying species boundaries in the network of genealogical relations, and at detecting barriers to gene flow between genetically distinct but sympatric groups (e.g. Lumbsch and Leavitt, 2011; Fujita et al., 2012).

The ant genus *Cataglyphis* comprises approximately 100 described species that inhabit arid land and deserts of the Old World, from Central Asia to North Africa (Lenoir et al., 2009). Although many phylogenetic studies based on morphology (Agosti, 1990; Radchenko, 2001), exocrine product chemistry (Dahbi et al., 1996, 2008; Keegans et al., 1992; Oldham et al., 1999), and molecular markers (Aron et al., 2016a; Knaden et al., 2012) have been conducted, the relationships among many *Cataglyphis* species remain ambiguous. This is particularly true for the species of the *bicolor* group, in which several studies have suggested the possibility that some of them contain several putative species across their wide geographic ranges, leaving an open question as to their phylogenetic relationships (Aron et al., 2016a; Knaden et al., 2012). For example, *C. savignyi,* currently classified as a single species, has been recorded from Morocco to Israel, *C. niger* is found from Israel to Yemen and *C. Isis* from Tunisia to Syria. The increase in morphology-based records of ants from previously unreported localities, together with recent empirical studies that focus on these species (Leniaud et al.,

* Corresponding author.
E-mail address: pieyer@live.fr (P.A. Eyer).

http://dx.doi.org/10.1016/j.ympev.2017.07.024
Received 9 January 2017; Received in revised form 22 June 2017; Accepted 26 July 2017
Available online 31 July 2017

2011; Saar et al., 2014) indicate that a more robust validation approach is required for the systematics of species in the *C. bicolor* group. The use of an integrative dataset incorporating both ancient branching events and current gene flow provides an opportunity to assess *C. bicolor* systematics in a rigorous statistical framework, while testing the hypothesis that species evolved independently as population-level lineages (de Queiroz, 2007; Schlick-Steiner et al., 2010). The current study, therefore, is not merely a taxonomic exercise, as achieving accurate estimates of *C. bicolor* species boundaries is necessary in order to study the transition zone between the Asian and African continents, as well as for extrapolating the evolutionary processes that drive the diversity of reproductive and social strategies within the genus *Cataglyphis* (Aron et al., 2016a, 2016b; Boulay et al., 2017). This group is appealing subject for this purpose as profound differences in social structure were found between populations across the relatively small geographic area of the Israel landscape, ranging from monogynous and monodomous colonies to highly polygynous, supercolonial populations (Leniaud et al., 2011; Saar et al., 2014). These differences in social structure may underlie the genetic differentiation between lineages, as they may create reproductive barriers between social forms or may reflect the social polymorphism of a unique species.

In order to resolve the phylogenetic relationships and study the cryptic species nested within the *C. bicolor* species complex (Aron et al., 2016a), we applied a comprehensive and integrative species delimitation approach to the analysis of 290 specimens collected from 84 localities in Israel. First, using nuclear and mitochondrial sequences we inferred divergence events across a range of evolutionary timescales and studied patterns of genetic variation in natural populations. Second, we compared this genetic-based phylogeny to classical phenotypic-based clustering using morphometric measurements and chemical analyses of cuticular hydrocarbons. Third, using microsatellite markers, we investigated current gene flow and pinpointed hybridization between sympatric but genetically distinct groups. Finally, we untangled the occurrence of cryptic species by applying four species delimitation approaches to our dataset. Overall, using these different sources of evidence we locate our samples within a global phylogeny of the *bicolor* group and discuss the processes that underlie speciation in this group.

2. Material and methods

Sampling comprised 290 individuals, each originating from a different nest and from 84 localities across Israel (Fig. 1a; Table S1; Supporting Information). As outgroup, we used two individuals of *C. altisquamis* (a *Cataglyphis* species that belongs to the *altisquamis* group) from two locations in northern Israel. All samples were stored in 98% ethanol until DNA extraction.

2.1. Morphometric analysis

The classical morphometric approach was adopted to identify species boundaries using 16 body measurements at a magnification of 50× to the nearest 0.01 mm, using a MZ8 stereomicroscope (Leica Microsystems, Wetzlar, Germany) on 495 individuals from 17 localities. Measurements comprised total length, head length and width, eye length, inter-ocular distance, ocellus-clypeal distance, scape length, mesosoma length and width, propodeum length and height, petiole node length, width and height, midfemur length, and hindtibia length (Ionescu and Eyer, 2016). *C. bicolor* species are characterized by a large gradual size polymorphism between workers (Knaden et al., 2005). To ensure robustness, we performed these measurements only on samples with more than eight ants per locality ($X \pm SD = 20.9 \pm 9.3$). To take into account the size of the individual, we constructed 15 covariable tests for each locality (Fig. S1; Supporting Information). The differences between each pair of localities were then tested using Tukey HSD post hoc tests. The results of these 15 tests between each pair of localities were summarized in a locality-by-locality matrix. In addition, a hypotheses-free ordination was built on our 16 morphological measurements using a Principal Coordinates Analyses using *FactoMineR* R package (Lê et al., 2008).

2.2. Chemical analyses

Chemical analyses of cuticular hydrocarbons were performed on samples from 27 colonies, selected on the basis of the molecular data as representing different clades in different localities of Israel. Each sample contains a pool of 10 heads from workers belonging to the same colony. They were extracted in 50 µl hexane (Sigma–Aldrich) and then evaporated to a volume of ca. 3 µl under a nitrogen stream prior to analysis. Analyses were conducted using coupled gas Chromatography mass spectrometry (GC/MS) with a VM-5 fused silica column that was temperature-programmed from 60 °C (1 min of initial hold) at 5 °C/min to 300 °C, with a final hold of 10 min. Compound identifications were inferred from their mass fragmentation and quantification by peak integration using ChemStation software (Agilent Technologies). The relationship between the examined species was assessed by a hypotheses-free ordination built on our 56 chemical compounds using a Principal Coordinates Analyses using *FactoMineR* R package (Lê et al., 2008).

2.3. Genetic analyses

Total genomic DNA was extracted from four legs of each individual using a standard Chelex-extraction protocol (Walsh et al., 1991). Nuclear and mitochondrial sequences, as well as microsatellite markers were selected to infer divergence events across a range of evolutionary timescales. Four nuclear loci (nu1281, 484 bp; nu1087, 593 bp; nu965, 374 bp; and Longwave Rhodopsin, LWRh, 656 bp), two mitochondrial loci (cytochrome *b*, CytB, 653 bp and cytochrome *C* oxidase, COI, 429 bp), and eleven microsatellite loci (cc26, cn04, cc51, cc99, cn02, cc54, ch11, ch23, ch08, ch01 and ch10) were examined for genetic analyses. All 292 individuals were sequenced for the CytB marker to infer lineages segregation on mtDNA. To confirm conclusions regarding mtDNA divergence, and to locate our samples within a global phylogeny of the *bicolor* group using GeneBank samples, a subset of 87 samples, representative of the mtDNA lineages uncovered using CytB marker (see Section 3), was analyzed for the COI marker. To study lineages segregation on nuclear DNA, 92 individuals were analyzed for the nu1281 marker, 79 for nu1087, 74 for nu965, 34 for LWRh, and 112 for the microsatellite genotyping. The nuclear genes studied are usually informative for phylogeographic inferences in ants (Brady et al., 2006); among them, three are Exon-Primed-Intron-Crossing (EPIC) markers to ensure a high polymorphism in the taxa investigated (Ströher et al., 2013). EPIC primer sequences were adapted from Ströher et al. (2013) to match *Cataglyphis* ants, while the microsatellite loci had been previously developed for *Cataglyphis* species (Pearcy et al., 2004; Clemencet et al., 2005; Darras et al., 2014). The PCR optimization for each locus, primer sequences, and bibliographic sources, as well as the method for detection of null-alleles and avoidance of linkage disequilibrium for microsatellite loci, are detailed in the Supporting Information (Table S2; Supporting Information).

For nuclear and mitochondrial sequences, PCR products were purified with the spin-column PCR purification kit (TIANGEN Biotech), followed by sequencing with the ABI BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, U.S.A.). Sequencing was performed on an ABI 3730 Genetic Analyzer (Applied Biosystems). Base calling and sequence editing were performed using CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA). For nuclear genes, this included assessment of gaps and reconstruction of parental haplotypes using Champuru software (Flot, 2007). Parental haplotypes were confirmed using PHASE with 0.90 thresholds and using known haplotypes (identified from homozygous individuals) as reference (Stephens et al., 2001); the PHASE input files were

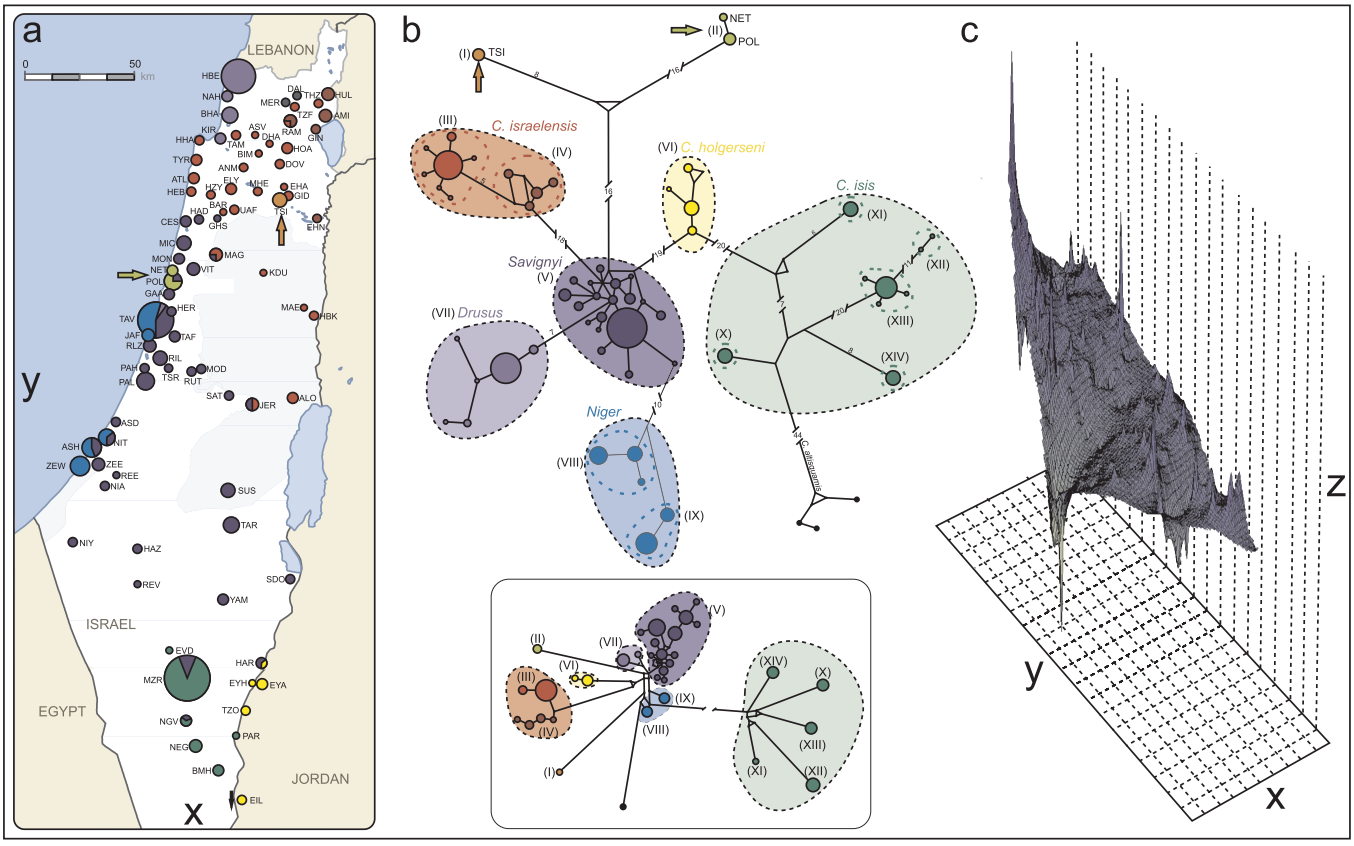

Fig. 1. (a) Sampling sites of *Cataglyphis* specimens studied in Israel. Each locality is colored according to its *mt*DNA lineage origin. Circle sizes are proportional to the number of individuals sampled. Dotted lines represent barriers to gene flow inferred on *mt*DNA dataset. (b) Haplotype network for the mitochondrial markers CytB and COI. Circle sizes are proportional to the number of haplotypes observed in the dataset and the branch lengths indicate the number of mutations between haplotypes. SAMOVA-identified groups are indicated with roman numbers. (c) Result of genetic landscape shape interpolation. Surface plot heights reflect genetic distance (and genetic diversity) patterns over the geographical landscape examined.

constructed using SeqPHASE (Flot, 2010). Sequences were aligned using MUSCLE algorithm (Edgar, 2004) implemented in CodonCode Aligner. All DNA sequences were deposited in the GenBank database (accession numbers are given in Table S1; Supporting Information). For microsatellite markers, fragment sizes were separated by an ABI 3730 Genetic Analyzer and determined using Gene Marker® HID (Holland and Parson, 2011).

2.4. Phylogeographic analyses

For both the mitochondrial and the nuclear dataset, phylogeographic relationships between haplotypes were represented on networks produced by the median-joining method (Bandelt et al., 1999) implemented in the program NETWORK v.4.6.1.1 (available at http://www.fluxus-engineering.com/).

For mitochondrial markers, significance of correlation coefficient between genetic differentiation and geographical distance was assessed with a Mantel test (Mantel, 1967), as implemented in Arlequin v.3.5.1.2 (Excoffier and Lischer, 2010). To detect geographical clustering of genetic variations within the Israeli landscape, we explored the patterns of genetic divergence among sampling localities using the SAMOVA (Spatial Analysis of Molecular Variance; Dupanloup et al., 2002) with SPADS v.1.0 (Dellicour and Mardulyn, 2014). The SAMOVA algorithm identifies the presence and the number of geographical groups that structure the genetic distribution. The presence of two to 84 (number of localities) genetic groups (K) was tested. The inter-group differentiation (ΦCT), which represents the part of the molecular variation explained by a K-group structure, was estimated for each value of K. The most likely number of genetic groups is expected by the maximal value of ΦCT. We also produced a 3D landscape-shape interpolation plot, where the x- and y- axes correspond to geographical locations and the z- axis represents genetic distances using Alleles In Space v.1.0 (AIS; Miller, 2005). The plot contains peaks in areas of large genetic differences, which can be used to identify landscape barriers to gene flow.

Comparisons within and between genetic groups were performed based on nucleotide diversity and genetic divergence using MEGA v. 5.0 (Tamura et al., 2011). The distribution of genetic variation was assessed using an Analysis of Molecular Variance (AMOVA) as implemented in Arlequin v. 3.5.1.2 (Excoffier and Lischer, 2010). Genetic variation was hierarchically partitioned among genetic groups (defined by SAMOVA analysis), among localities within groups, and among individuals within localities. The variance components were tested statistically by non-parametric randomization tests of 10,000 permutations of the original data.

We used 11 microsatellite loci on a subset of 112 samples to further investigate genetic diversity and the occurrence of gene flow between genetically distinct but sympatric groups. The number of genetic clusters or subpopulations (K) in our sampling was assessed using Bayesian clustering approaches implemented in Structure v.2.3 (Pritchard et al., 2000). The simulations were run with values of K from 1 to 50 and repeated 20 times for each number of K. A combination of an admixture and a correlated-allele frequencies model was used for the analysis. Each run included a 5×10^4 burn-in period followed by 1×10^5 iterations of the MCMC. The most likely number of groupings was evaluated using ΔK (Evanno et al., 2005) and log-likelihood value by Structure Harvester v.0.6.8 (Earl and vonHoldt, 2012).

2.5. Gene trees

Maximum Likelihood (ML) and Bayesian Inference (BI) were used to reconstruct phylogenetic inferences for the two mitochondrial and four nuclear sequence markers. The best fitting nucleotide models for each gene were selected on the basis of minimal value of the Bayesian information criterion (BIC) in order to avoid overparametrization. We explored data partitioning that would allow each codon position of each gene to have its own substitution model. We calculated the best partitioning scheme and its substitution models using PartitionFinder v.1.0.1 (Lanfear et al., 2012). We investigated three different partitions for each locus (1st, 2nd and 3rd codon positions). ML analyses were conducted with the PhyML online web server v.3.0 (Guindon and Gascuel, 2003; Guindon et al., 2010). Nodal support was assessed by bootstrap resampling analysis (1000 pseudoreplicates). BI was implemented with the program MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003) under a partitioned model considering nucleotide substitution models estimated with PartitionFinder v.1.0.1. The Bayesian posterior probabilities were estimated using a Metropolis-coupled Markov chain Monte Carlo sampling approach, with two runs starting with random trees running for $4x10^6$ generations, saving one tree every 100 generations to produce 40,000 trees. Both convergence and appropriate sampling were confirmed by examining the standard deviation of the split frequencies between the two simultaneous runs (< 0.01) and the Potential Scale Reduction Factor (PSRF) diagnostic. The first 10,000 trees of each run were included in the burn-in period and discarded, and a majority-rule consensus tree was generated from the remaining trees.

2.6. Species delimitation approaches

We examined the presence of cryptic species within our sampled individuals on the basis of four different species delimitation methods: the Generalized Mixed Yule-Coalescent model (GMYC; Pons et al., 2006), the Automatic Barcode Gap Discovery (ABGD; Puillandre et al., 2012), the BP & P (Rannala and Yang, 2003; Yang and Rannala, 2010), and the Haploweb (Flot et al., 2010). These methods enable the delimitation of independently-evolving species based on genetic data (*e.g.,* Lahaye et al., 2008; reviewed in Fujita et al., 2012; Dellicour and Flot, 2015) and do not require *a priori* hypotheses of putative species groupings (based on taxonomy, morphospecies, or chemistry), thereby limiting potential bias in species delimitation. We selected these four different delimitation methods as they use quite dissimilar approaches to delimiting species.

The ABGD method is a single-gene approach to the statistical detection of barcode gaps in a pairwise genetic distance distribution (Puillandre et al., 2012). Barcode gaps, which presumably occur between intraspecific and interspecific distances, were used to partition our mitochondrial dataset into species hypotheses (initial partition). The resulting inferences were then recursively applied to yield finer partitions (recursive partitions) until no further partitioning was possible. ABGD analysis was run on the mitochondrial CytB and COI markers through a web-based interface (http://www.abi.snv.jussieu.fr/public/abgd/) using the default parameters (10 steps of intraspecific divergence prior from $P\text{min} = 0.001$ to $P\text{max} = 0.10$, $X = 2$).

The GMYC method is also a single-gene approach to identifying species 'boundaries', which are associated with shifts in branching rates between intra- and interspecies branching events on a time-calibrated ultrametric tree (Pons et al., 2006; Fujisawa and Barraclough, 2013). Branching patterns within species reflect neutral coalescence processes, whereas branching patterns among species reflect the timing of speciation events. This method exploits the predicted difference in branching rate under the two modes of lineage evolution, assessing the point of highest likelihood of the transition (i.e., threshold; Pons et al., 2006; Zhang et al., 2013). We used a Bayesian implementation of this method (bGMYC; Reid and Carstens, 2012), which allows single-threshold algorithms and improves delimitation by means of more exhaustive searching based on Bayesian Markov Chains Monte Carlo. The bGMYC was run in an R package (R Development Core Team, 2010; Reid and Carstens, 2012) using 100 ultrametric gene trees reconstructed under a relaxed molecular clock model (uncorrelated lognormal), using BEAST v.1.7.4 (Drummond et al., 2012). The BEAST input files were generated using BEAUti v.1.7.4 (Drummond et al., 2012), with both the mitochondrial CytB and COI markers. MCMC chains were run over 50 million generations and a tree was sampled every 1000 generations. Tree congruence across runs and Effective Sample Sizes (ESSs) were visually confirmed in Tracer v.1.7.4. The first 25% of trees were then discarded as a burn-in and 100 trees were randomly selected among the remaining trees to run in the bGMYC package. The bGMYC analysis ran with the followed parameters: 500,000 MCMC generations, 100 thinning, and 10% burn-in.

In contrast to ABGD and bGMYC, BPP is a multi-locus approach. This Bayesian species delimitation was conducted on the full phased dataset of the four nuclear loci and the two mitochondrial genes using BPP v2.2 (Rannala and Yang, 2003; Yang and Rannala, 2010). The model assumes that different genes may be in conflict regarding the species tree, and hence a single gene tree may not mirror the speciation process. This is particularly the case for recently diverging groups for which monophyly has not been reached due to incomplete lineage sorting (Hudson and Coyne, 2002). In these groups, the criteria of reciprocal monophyly and strict congruence between gene trees could be too strict to enable detection of putative species boundaries. BPP calculates posterior probabilities of potential species delimitations using a user-specific-species-tree. The BPP analysis was run for 200,000 generations with a sampling interval of 1 and a 20% burn-in, automatically adjusted fine-tuning of the parameters and with an equal prior probability assigned to each species delimitation model. The species tree obtained with *BEAST was used to assess the candidate species (see below).

Contrasting the other approaches, the Haploweb method is based on diploid phased nuclear markers and is particularly suited to delineate species in species-poor datasets (Dellicour and Flot, 2015). It consists in linking haplotypes (alleles) that co-occur within heterozygous individuals on median-joining haplotype networks (Flot et al., 2010). The allele pool composed by the interconnected haplotypes is called a "field for recombination" (FFR), i.e., a putative species (Doyle, 1995).

2.7. Species tree

The species tree and divergence times were calculated using a coalescence-based approach as implemented in *BEAST (Heled and Drummond, 2010). After excluding potential hybrids or introgressed populations (i.e. POL and TSI, in order not to violate the species tree model assumptions), we chose to use each *mt*DNA lineage as a candidate species, and therefore applied a six-species model. Both mitochondrial fragments and the four unlinked phased nuclear alleles were used in the *BEAST analyses, resulting in six independent partitions. To replace our samples in the global phylogeny of the *C. bicolor* group, we also performed a BEAST analysis while adding the COI samples belonging to the group *C. bicolor* available in GenBank to a subset of our samples (Table S1, Supplementary Information). Each *BEAST and BEAST analysis was performed with two independent runs of 20×10^7 generations, sampling every 10,000 generations, from which 10% were discarded as burn-in. Models and prior specification were as follows (otherwise by default): GTR + I + G model for both mitochondrial COI and CytB, and HKY + I + G for the four nuclear markers; Relaxed Uncorrelated Lognormal Clock (estimated for nuclear genes, 2.3% My^{-1} for mitochondrial genes; Papadopoulou et al., 2010); Yule process of speciation; random starting tree; alpha Uniform (0, 10). Convergence for all model parameters was assessed by examining trace plots and histograms in Tracer v1.7.4 (Drummond et al., 2012) after obtaining an effective sample size (ESS) > 200. Runs were combined

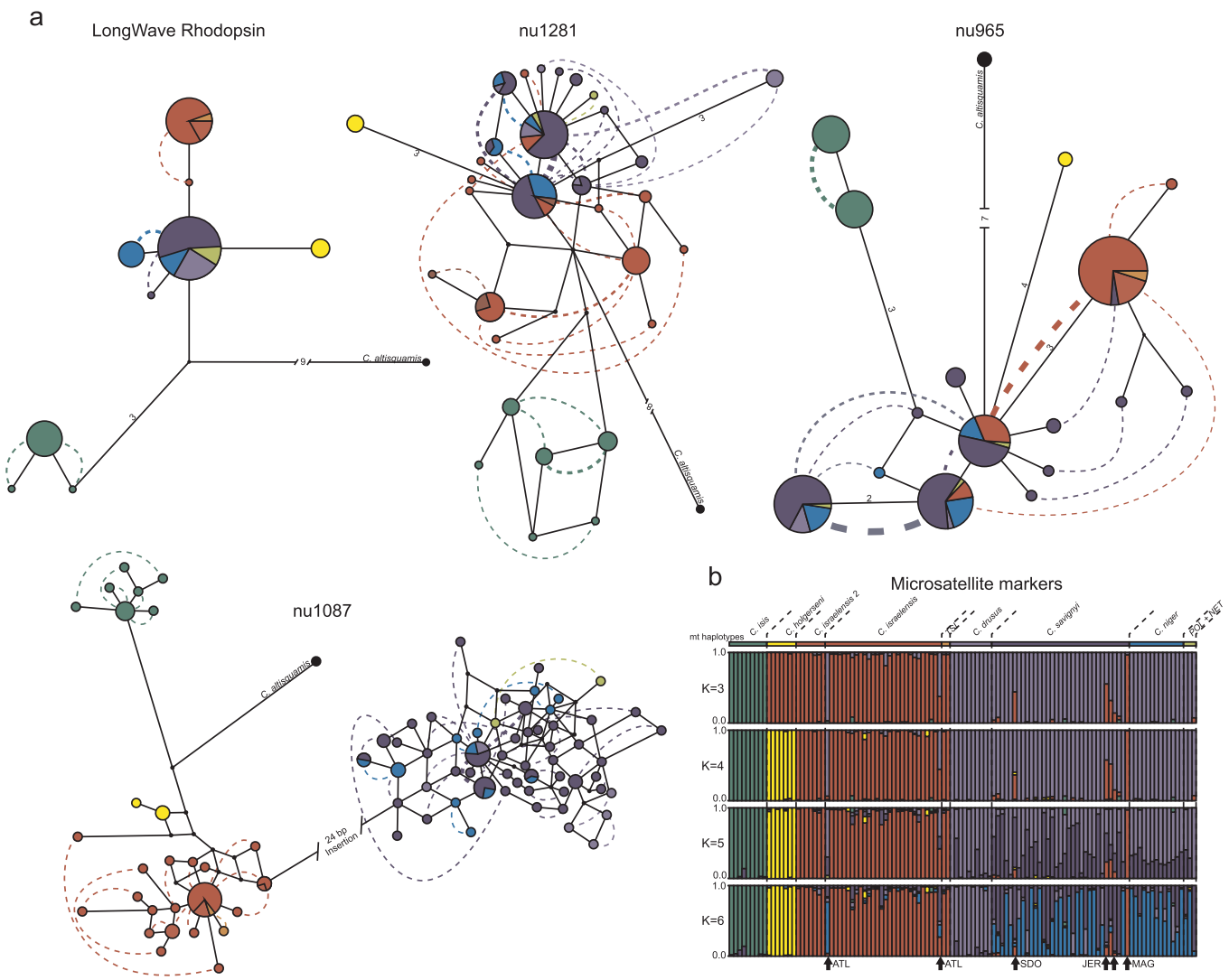

Fig. 2. (a) Haplotype webs (haplowebs) for the four nuclear genes studied. Circle sizes are proportional to the number of haplotypes observed in the dataset and colored according to their *mt*DNA origin. Branch lengths indicate the number of mutations between haplotypes. Dashed lines between two haplotypes represent heterozygous haplotypes co-occurring in the same individual and their thickness is proportional to the number of heterozygous individuals. (b) Graphical representation for Structure results for different values of K genetic groups. Each group is characterized by a color; and each individual is represented by a vertical bar according to its probability to belong to each group. Arrows indicated potential hybrids between groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using LogCombiner, and maximum credibility trees with divergence time means and 95% highest probability densities (HPDs) were produced using Tree Annotator (both part of the BEAST package). Trees were visualized using FigTree v1.3.1 (available at http://tree.bio.ed.ac.uk/software/figtree).

3. Results

3.1. Phylogeogarphic analyses

3.1.1. Mitochondrial DNA

Overall, we examined 292 individuals at the CytB mitochondrial marker resulting in 429 bp sequences with 131 variable nucleotides (124 PI), and identified 60 unique haplotypes. We also analyzed a subset of 87 individuals at the COI mitochondrial marker (653 bp, 146 variable nucleotides, 101 PI and 35 haplotypes). Both mitochondrial networks recovered strong geographical structuring of diversity with at least eight deeply divergent lineages (Fig. 1b). This mitochondrial structure was reinforced by SAMOVA analysis, which further split the *mt*DNA divergence and determined K = 14 as a probable number of genetic groups within our sampling (Fig. 1b). The net sequence divergence between the observed lineages was rather variable, ranging from 14.63% between the *C. isis* and the *Niger* lineages, to 2.65% between the *C. drusus* and the *C. savignyi* lineages (Table S3). Geographically, the mitochondrial lineage corresponding to *C. isis* samples spans the southern Negev, while *C. holgerseni* haplotypes form another lineage located only in the Arava valley (Fig. 1a). Despite the small size of the sampling area, *C. isis mt*DNA haplotypes revealed an extensive within group diversity (6.80%), whereas *C. holgerseni* displayed rather uniform haplotypes (0.47%) for a similar sampling area size (Table S3). Quite surprisingly, a *new* lineage was discovered, predominantly in the mountainous areas from Jerusalem to the Golan Heights, although its distribution also extends to the coastal plains in Atlit where Mount Carmel reaches the Mediterranean Sea. This lineage is hereafter named *C. israelensis*, according to its recent description (Ionescu and Eyer, 2016). The sequence divergence between *C. israelensis* and the other lineages ranged from 13.12% (*C. isis*) to 6.64% (*Savignyi* lineage) (7.88% between *C. israelensis* and *C. holgerseni*). This lineage shows rather weak within-group diversity (0.75%), despite some substructure with various haplotypes occurring in its northern distribution. The *C. savignyi* lineage is the most widespread in Israel, from the Negev to the coastal plain. Despite its large distribution range, *C. savignyi mt*DNA

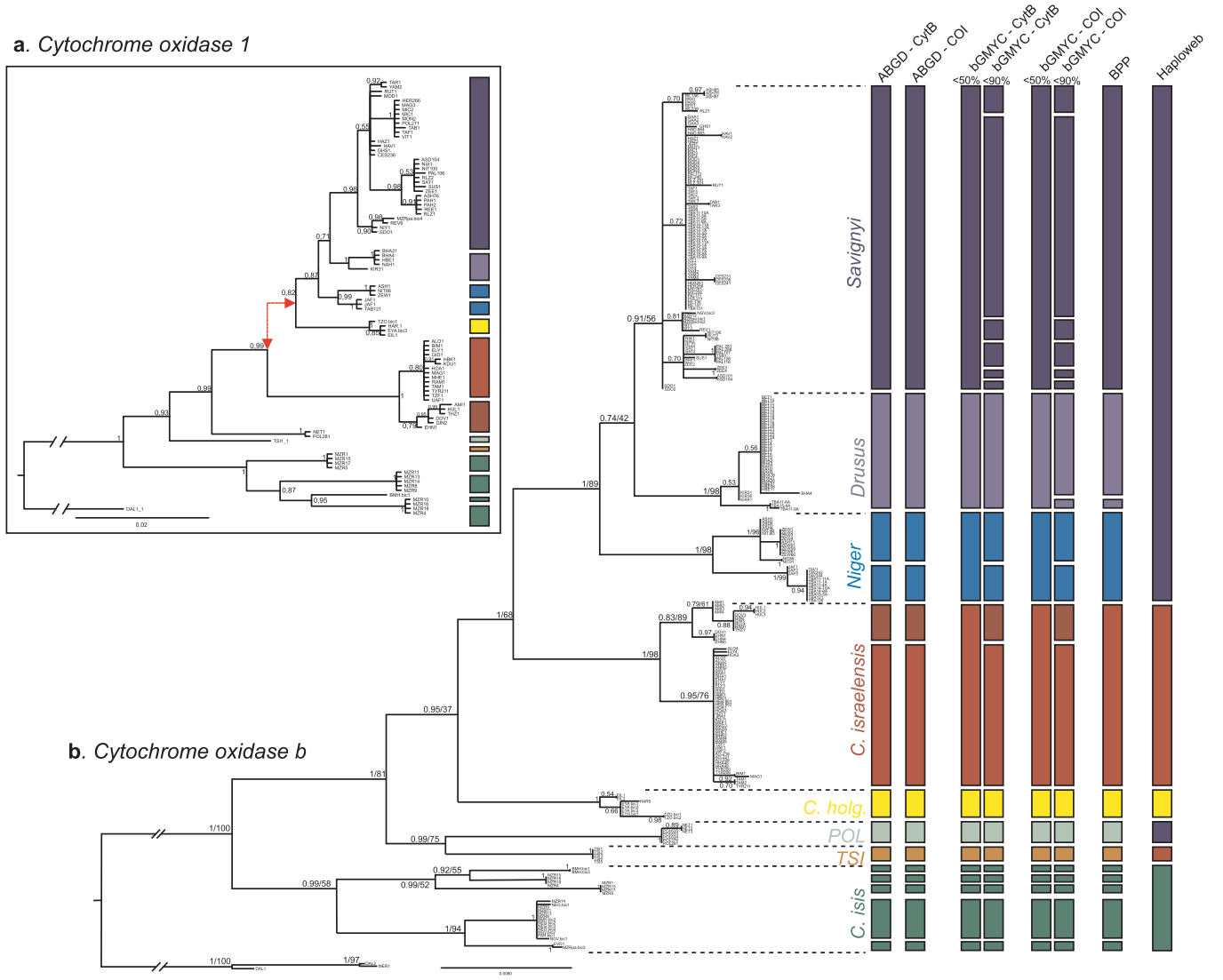

Fig. 3. Bayesian inference trees of (a) 87 COI haplotypes and (b) 292 CytB haplotypes of *Cataglyphis* ants. Bayesian probabilities/bootstrap values (from ML tree) are given to estimate branches support. The trees are rooted using the haplotypes of *C. altisquamis* as outgroup. Results of species delimitation analyses using ABGD, bGMYC, BPP and Haploweb methods are indicated in front of each sample.

haplotypes do not display a high diversity (0.50%). Two *mt*DNA lineages occur in restricted distribution: the *C. niger* lineage in the coastal plain of the Tel Aviv area and north of the Gaza Strip, and the *C. drusus* lineage in the northern tip of the Israeli coast, north of Mount Carmel. Quite surprisingly, two distant sets of haplotypes were found in two unique vicinities, (POL and NET) on the central coastal plain and (TSI) in Lower Galilee. Overall, we found a substantial and significant *mt*DNA differentiation among populations (F_{ST} = 0.822; P < 0.001), and a significant relationship between geographical and genetic distances (Mantel test, P < 0.001). Landscape shape interpolation analysis produced a 3D surface plot, showing peaks in areas where there are large genetic differences (i.e. landscape barriers to gene flow). In this plot, differentiation increased from largely smooth (low barriers to gene flow) genetic surfaces in the central Negev (only *C. savignyi* lineage) or the Galilee (only *C. israelensis* lineage) towards regions of elevated genetic distance in coastal areas and the southern Negev (Fig. 1c). This roughly corresponds to the high genetic diversity revealed along the coast where the *C. savignyi*, *C. niger*, and *C. drusus* lineages co-occur, and in the sympatric region of the *C. isis*, *C. holgerseni*, and *C. savignyi* lineages in the southern Negev and Arava.

3.1.2. Nuclear DNA

Overall, after splitting the heterozygous alleles for the four nuclear markers used, our nuclear dataset has resulted in 39 sequences analyzed from LWRh, 145 from nu1281, 135 from nu1087, and 115 from nu965. Nine unique haplotypes were revealed from LWRh, 31 from nu1281, 87 from nu1087, and 16 from nu965 (Fig. 2a).

Due to the presence of a high number of gaps in the EPIC markers, only the network analyses are shown, while the BI and ML trees are given in Supporting Information (Fig. S2). The networks reveal a well-defined separation of *C. isis* and *C. holgerseni* from the other *C. bicolor* species on all four nuclear genes. To a lesser extent, the nuclear clustering confirms the separation of the *C. israelensis* lineage previously revealed by the *mt*DNA analysis. Nuclear haplotypes of *C. israelensis* are clearly distinct on the LWRh and nu1087 markers, but separation is barely visualized on both nu1281 and nu965, as some haplotypes are shared between the *mt*DNA lineages (Fig. 2a). In contrast, the situation for *C. drusus*, *C. niger*, and *C. savignyi mt*DNA lineages is more problematic as they are indistinguishable on any of the nuclear markers. Interestingly, haplotypes from the highly divergent mitochondrial lineages POL-NET and TSI are integrated in the cluster comprising *C. drusus, C. niger, and C. savignyi* lineages (hereafter named *C. niger* complex) and in the *C. israelensis* cluster, respectively.

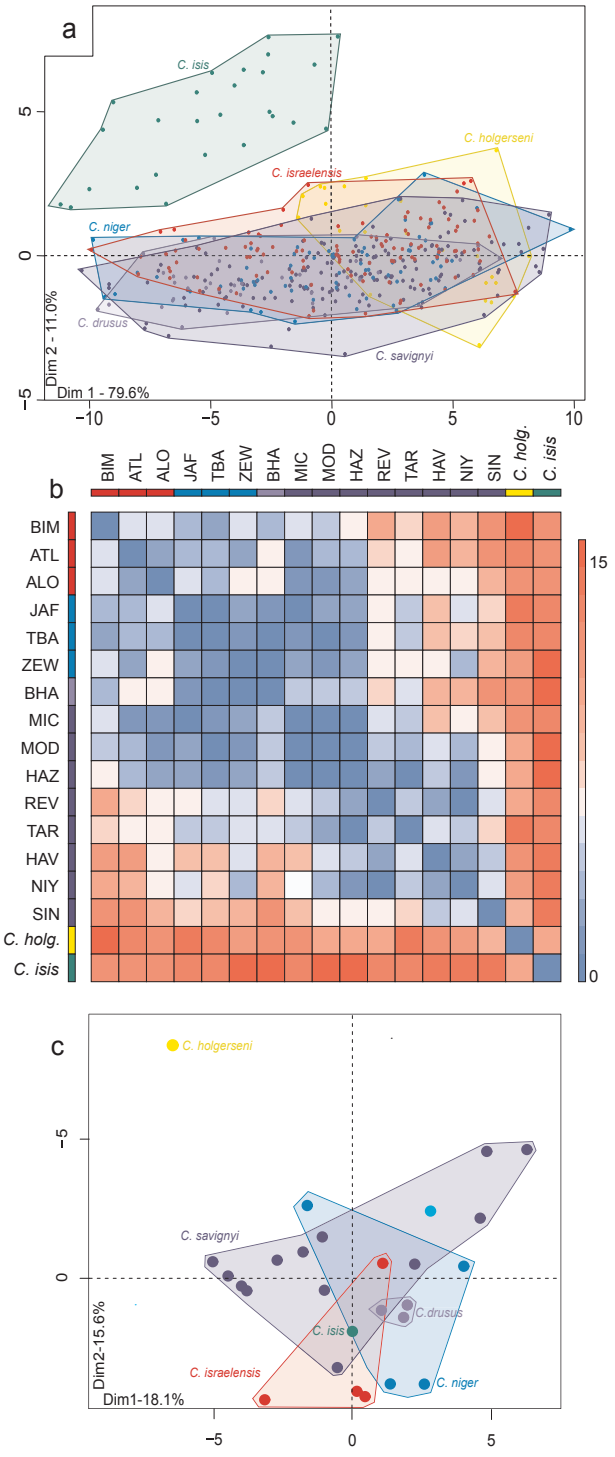

Fig. 4. Results of morphology-based (a & b) and chemical-based (c) species delimitation. (a) Principal Component Analysis of 16 morphological measurements from 495 samples of the *C. bicolor* group. (b) Results of the 15 co-variables analyses are summarized using a colored population-by-population matrix (they are detailed in Fig. S1). Cells are colored according to the number of significant co-variables test between each pair of populations. (c) Principal Component Analysis of 56 chemical compounds from 27 colonies of the *C. bicolor* group. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.3. Microsatellite markers

Microsatellite marker analysis by Structure indicated the occurrence of three distinct genetic groups in our sampling, following the ΔK method developed by Evanno et al. (2005) (Fig. 2b). Accordingly, *C. isis* individuals were grouped together as a single cluster; *C. holgerseni* was surprisingly grouped with *C. israelensis* in the second cluster; and, similar to the nuclear networks, *C. drusus*, *C. niger*, and *C. savignyi* were grouped as a single cluster. Increasing the number of K allows the genetic distinctiveness of *C. holgerseni*; however *C. drusus*, *C. niger*, and *C. savignyi* *mt*DNA lineages remain indistinguishable irrespective of the number of K investigated. Consistent with the results for the nuclear markers, individuals with highly divergent mitochondrial haplotypes from POL-NET localities are integrated in the cluster comprising the surrounding *C. savignyi, C. drusus,* and *C. niger* samples, while the TSI samples are not separated from those of *C. israelensis*. Noticeably, several hybrids between the *C. israelensis* and *C. savignyi* lineages were found in localities at the border of their distribution ranges and in localities where both lineages are sympatric (i.e., MAG, ATL, SDO & JER). Some of these hybrids are F1 hybrids (50% admixture), but also backcrossed individuals (< 50%) or introgressed individuals (near 100% of microsatellite assignment from one lineage but bearing the *mt*DNA haplotype from the other one).

3.2. Phylogenetic reconstruction and species delimitation

ML and BI methods led to similar tree topologies on mitochondrial markers COI and CytB (Fig. 3a and b); hence, only trees based on BI are presented here (ML trees are given in Fig. S3). They both highlight the same 8–15 lineages recovered in the mitochondrial networks and the SAMOVA analyses.

The full results of the bGMYC, ABGD, and BP & P species delimitation analyses for the COI, CytB, and combined *mt/nu*DNA datasets are presented in the Supplementary Information (Figs. S4 and S5). A summary of these results is represented on the mitochondrial trees (Fig. 3b). For bGMYC, *Cataglyphis* specimens were assigned to from 9 (posterior probability: $0.5 < P < 0.9$) to 14 ($P > 0.9$ level) taxa for the CytB, and from 12 to 20 for the COI (Fig. 3b and S4). Consistent with these results, the ABGD method detected 14 clusters in our samples using both mitochondrial datasets (Fig. 3b and S5). The grouping of the *Cataglyphis* samples was quite congruent between the ABGD and GMYC analyses on both the CytB and the COI dataset. The clusters revealed by these two methods correspond to the groups highlighted by the SAMOVA analysis. Multiple runs of BP & P with different MCMC parameters and guide trees gave similar results, with all internal nodes having high speciation posterior probabilities (SPP > 0.99). Regardless of the prior distributions used, BP & P results show that all nodes separating the three groups of recently diverging sequences of *C. niger*, *C. drusus*, and *C. savignyi* are supported by very high speciation posterior probabilities. Haploweb analyses of each nuclear marker yielded four (LWRh), three (nu1281), three (nu965) and four (nu1087) fields for recombination (FFRs), i.e., putative species (Figs. 2a and 3b) among the studied samples. The four putative species identified by the Haploweb approach are congruent with the four groups revealed by Structure analysis using microsatellite markers: *C. isis*, *C. holgerseni*, *C. israelensis*, and the *C. niger* complex. The separation between *C. israelensis* and the *C. niger* complex is clearly supported on LWRh and nu1087 loci, but is barely noticeable on nu965 and nu1281 as some haplotypes are shared between the two genetic groups. However, the three lineages *C. drusus*, *C. niger*, and *C. savignyi* were not distinguishable on any of the nuclear loci analyzed. Overall, the two single-gene species delimitation methods based on the mitochondrial datasets (ABGD & bGMYC) tended to over-split groups in comparison to the multilocus approaches (BPP & Haploweb analyses) (Fig. 3b).

3.3. Morphometric and chemical analyses

Approximately 90% of the morphometric variation could be summarized in the first two principal components. Morphometric analyses revealed strong dissimilarity between individuals (Fig. S1). Only *C. isis* is clearly distinct with no overlap, while *C. holgerseni* barely cluster apart from the remaining samples (i.e., *C. israelensis*, *C. savignyi*, *C.*

drusus and *C. niger*; Fig. 4a). A similar result is noticeable using co-variable tests (Fig. 4b), where *C. isis* and *C. holgerseni* are well distinguishable from each other, with 10 co-variable tests differing significantly between them. They are also morphologically distinct from the other samples examined on 9–15 (out to 15) co-variable tests (Fig. S1). Almost no difference occurs between *C. israelensis*, *C. savignyi*, *C. drusus* and *C. niger* samples. Only weak morphological differences were found using co-variable tests (X ± SD = 4.23 ± 3.15), and mainly occurred between the southern Negev and northern localities. Within the northern localities no major differences were found between the *C. israelensis*, *C. savignyi*, *C. drusus*, and *C. niger* localities (X ± SD = 2.28 ± 2.03 out of 15 co-variable tests).

Cuticular hydrocarbon analyses revealed chemical differences among the 27 colonies analyzed, with colonies from the same species roughly clustering together (Fig. 4c, Table S4). The main difference was found between *C. holgerseni* and the remaining species. Quite surprisingly, *C. isis* clusters with *C. savignyi*, *C. drusus*, and *C. niger*, while well separated on genetic markers.

3.4. Species tree

Our nuclear and mitochondrial sequences were combined in *BEAST to estimate the species tree for the six main different mitochondrial lineages found in Israel and revealed in the network and species delimitation analyses (POL-NET and TSI samples were removed from the analyses; Fig. 5a). We also used BEAST analysis on the COI marker to estimate time of divergence among all *C. bicolor* species (including our POL-NET and TSI samples; Figs. 5b and S6 for the non-collapsed tree). Molecular clock calibration establishes the origin of the *C. bicolor* clade at 3.02 millions years ago (Mya) (95% Confidence Intervals, CI = 2.2–4.0), with the divergence of *C. bicolor* from *C. altisquamis* occurring 11.22 Mya (CI = 17.5–7.0) (Fig. 5b). Lineages within *C. bicolor* began to diverge in the late Pliocene between 3.02 and 2.37 Mya (CI = 1.6–3.1), with an initial divergence of African (*C. viatica* and *C. bicolor*) and Asian (*C. longipedem* and *C. bellicosa*) species from the rest of the group (Fig. 5). The *C. isis* species, which form a unique, highly diverse clade from Syria to Morocco separated at 2.09 Mya (CI = 1.5–2.8). The *C. bicolor* clade from the northern Middle East (i.e., Turkey, Greece, and Lebanon, and our introgressed sequences from POL, NET and TSI) diverged from the southern Middle East (i.e., Israel, Jordan, Yemen and Oman, including one clade from Africa) at 1.38 (CI = 0.9–1.7) Mya. As expected, the mitochondrial lineages *C. niger, C. drusus*, and *C. savignyi* recently split from each other, approximately 0.57 Mya (CI = 0.4–0.8).

4. Discussion

The results of our study provide new insights into the systematics of *Cataglyphis* species in the *bicolor* complex in Israel, supporting the occurrence of at least four distinct species, one of which (the *C. niger* complex) may comprise three more recent species (specifically discussed below). First, we confirm the presence and distinctiveness of *C. isis* and *C. holgerseni* in the southern Negev. These two species were clearly separated from the others on all genetic markers tested: mitochondrial DNA and both nuclear sequences and microsatellite markers. This separation is also discerned in the morphological measurements and partially also in the chemical analyses. Second, our data reveal the presence of a strongly supported genetic clade, on several mitochondrial and nuclear sequence markers, in the central mountain ridge in Israel with a limited extension to the northern coastal plain. This clade probably represents the cryptic species *C. israelensis,* barely distinguishable on morphology from the other *C. bicolor* species in the region (Ionescu and Eyer, 2016). This clade is, however, chemically distinct from the *C. niger* complex, including the *C. niger*, *C. drusus*, and *C. savignyi* samples. Clustering analysis based on microsatellite markers mostly supports the integrity of this genetic group, suggesting a reduced gene flow between individuals belonging to this clade and those from other *C. bicolor* clades, possibly due to reproductive isolation. Notwithstanding, in some contact zones, rare hybrids (3%) were genotyped, suggesting an incomplete reproductive isolation. Third, our data reveal the occurrence of another genetic clade distributed from the northern Negev to the Mediterranean coast, including the previously described species *C. niger* and *C. savignyi,* and a third population that tentatively belongs to *C. drusus*. This clade is genetically distinct from the other species (*C. isis*, *C. holgerseni* and *C. israelensis*) on almost all genetic markers. Within this clade, our mitochondrial data reveal three geographically structured subgroups corresponding to *C. niger*, *C. savignyi* and *C. drusus* populations. However, several points cast doubt on their classification as distinct species: (i) this regional subgrouping is not supported by any of the nuclear sequences; (ii) some *C. savignyi* and *C. drusus* mitochondrial haplotypes also occur within the *C. niger* population; and (iii) microsatellite analysis revealed no genetic distinctiveness among these three populations, suggesting a current gene flow among them. Possible explanations for this genetic structure are either a recent speciation or a geographical structuring of a single species.

4.1. Species delimitation

The occurrence of at least four distinct *Cataglyphis* species of the group *bicolor* in Israel is supported by several findings. The two single-gene approaches based on *mt*DNA (bGMYC and ABGD) and the combined nuclear and *mt*DNA approach (BP & P) recognize at least the four major clades (but see below for other clades revealed by these analyses). These four clades are also reported in sequences from multiple independent nuclear loci (Haplowebs) and Structure analysis based on microsatellite dataset.

In the southern Negev, *C. isis* reveals a remarkable within-species *mt*DNA diversity, which exceeds the interspecific divergence between *C. bicolor* species, suggesting the occurrence of five cryptic species within *C. isis*. However, this interpretation is not supported by any of the nuclear genes, which cluster all the *C. isis* samples as a unique independently evolving lineage, i.e., a putative species. Interestingly, our samples of *C. isis* form a single *mt*DNA clade integrating those of *C. isis* and *C. diehlii* from GenBank (Figs. 5 and S6); and the *mt*DNA diversity revealed in our samples in Israel is higher than that found in the GenBank samples of *C. isis* and *C. diehlii* from a wide geographic range, including Syria, Israel, Tunisia, and Morocco (Knaden et al., 2012). This suggests that *C. isis* and *C. diehlii* may form a single species with a high mitochondrial diversity from Syria to Morocco.

The results for *C. holgerseni* indicate a distinct species based on all the studied genetic, chemical and morphometric characters. This species displays a weak genetic diversity despite its wide distribution range, a finding corroborated by the sequence similarity between the Israeli (this study) and the Yemen samples of *C. holgerseni* (which was published as *C. niger;* Knaden et al., 2012) (Figs. 5 and S6).

Our mitochondrial data reveal the presence of a clearly distinct lineage in the central mountain ridge in Israel, which represents the recently described species *C. israelensis.* This clade is supported by two nuclear genes, but its distinctiveness from the *C. niger* complex is hardly discernible when employing the two other nuclear genes, due to some shared haplotypes between the two genetic groups. Such conflicting trees may be expected in newly-formed species, since a substantial period of time is required following species divergence, in order to reach monophyly, due to incomplete lineage sorting or insufficient accumulation of mutations (Carstens and Knowles, 2007; Hudson and Coyne, 2002). The microsatellite analysis, which revealed that this lineage is genetically isolated from the other *C. bicolor* species studied, supports the conclusion that this group is an independently evolving lineage with restricted gene flow with other clades, possibly due to reproductive isolation. Notwithstanding, the rare occurrence of hybrid individuals (3%) in sympatric zones with the *C. niger* complex, may indicate an incomplete reproductive isolation. Finally, no match in

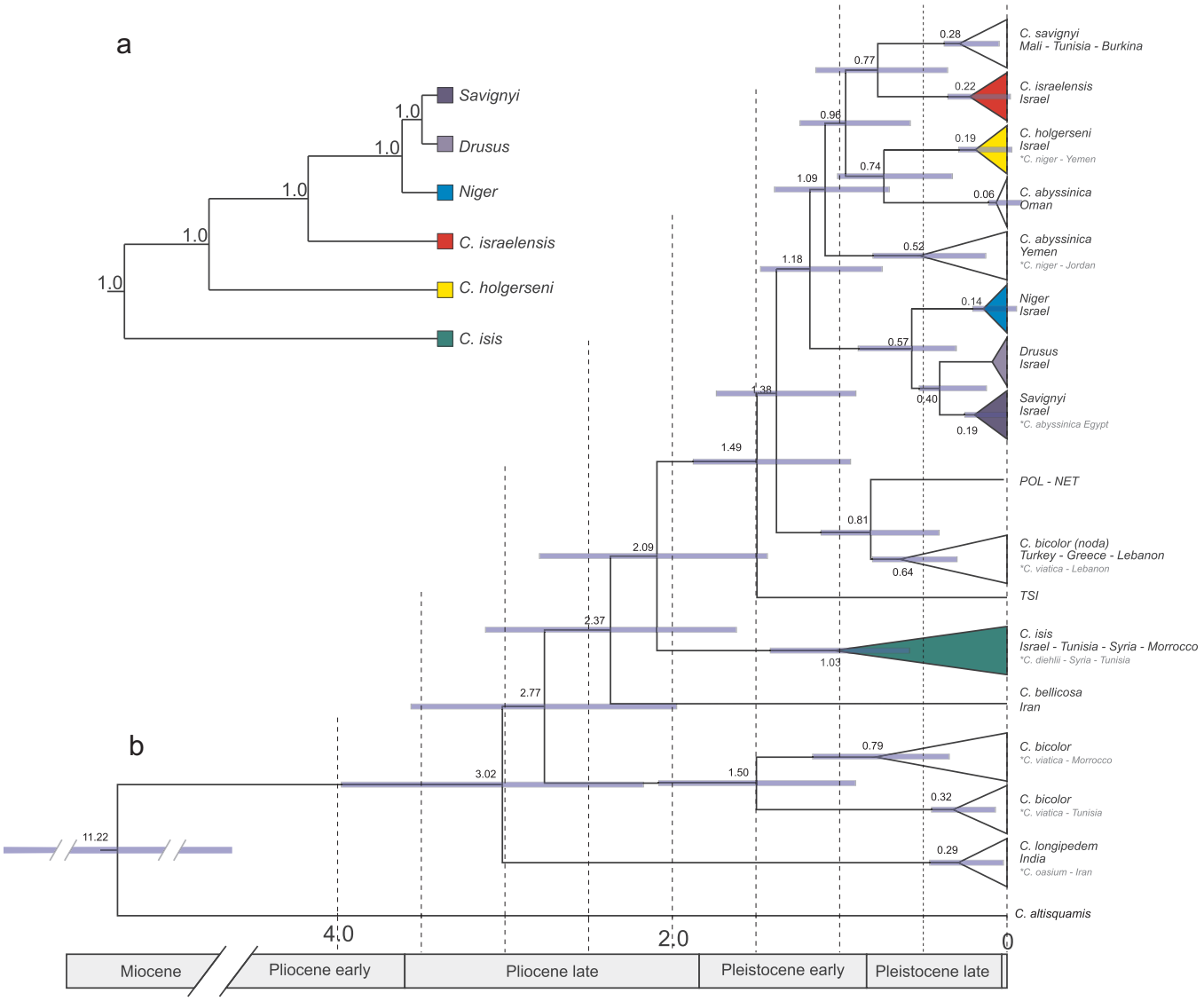

Fig. 5. (a) Topologies of Israeli *C. bicolor* lineages analyzed with multispecies coalescent models in *BEAST. The analyses are based on sequences of the 2 mitochondrial and the 4 unlinked phased nuclear markers. Branch labels indicate posterior probabilities for each lineages using multilocus species delimitation in BPP software. (b) BEAST tree of all bicolor species available in GenBank. Branch tips were collapsed to higher taxa. Asterisks indicate alternative or mislabeled names found in GenBank. Uncollasped tree is given in supporting information (Fig. S6).

*mt*DNA was found in GenBank for the data on *C. bicolor*, strengthening the conclusion that this clade represents a previously undescribed cryptic species, formerly merged with species of the *C. niger* complex (Vonshak and Ionescu-Hirsch, 2009).

The fourth species revealed with confidence by our study (i.e., the *C. niger* complex) is distributed from the northern Negev to the Mediterranean coast and encompasses the previously described species *C. niger* and *C. savignyi* and the putative *C. drusus* (discussed below). Although the mitochondrial data reflect the delimitation of the above three species, neither the nuclear sequences nor the microsatellite analysis support this. It is important to note that the *C. drusus* population was somewhat more homogenous on the microsatellite markers compared to *C. niger* and *C. savignyi* at a higher number of K (Fig. 2b, K = 6), which may derive from the separation of the *C. drusus* population from the C. *savignyi*/*C. niger* populations both by Mount Carmel and the invasion of *C. israelensis* into the coastal plain.

The discrepancy in species delimitation between the mtDNA and nuclear gene data reflects the notion that unique gene topologies do not mirror the speciation process (Carstens and Knowles, 2007; Heled and Drummond, 2010; Taylor et al., 2000). Two hypotheses may explain the lack of congruency between the *mt*DNA and nuclear gene lineages. The first posits a single species without reproductive isolation between the *mt*DNA lineages that allows continuous between-populations gene flow (due to high dispersal of males), while still maintaining a mitochondrial geographical substructure resulting from a low dispersal of newly-mated queens. The second hypothesis posits that the three populations represent distinct species, but that speciation is still too recent to have allowed lineage sorting within species on nuclear genes (Sullivan et al., 2002; Carstens and Knowles, 2007). However, faster-evolving microsatellite markers did not cluster into the three genetic lineages either, indicating the occurrence of current gene flow between them. It is worthwhile noting, however, that in the case of recent species, it might be impossible to differentiate between current hybridization and incomplete lineage sorting, as both of these processes are expected to occur in recent species with incomplete reproductive barriers and to overlap to produce similar genetic signatures (Holder et al., 2001; Holland et al., 2008; Joly et al., 2009).

Interestingly, earlier studies with some of these populations revealed remarkable differences in their social structure as well as in their mating systems. Inland populations of *C. savignyi* in the Negev (i.e. Arad

park) are monogynous and monodomous (each colony has a single queen and a single nest; Leniaud et al., 2011); while nearer the seashore, the populations of *C. savignyi* (i.e. Palmahim and Rishon Leziyon) are monogynous and polydomous (colony with single queen and multiple nests; Saar et al., 2014, published under *C. niger* name). In contrast, the *C. niger* population, which occurs in a similar habitat (i.e., Ashqelon), is highly polygynous (multiple queens per colony) and organized in supercolonial populations composed of numerous interconnected nests (Leniaud et al., 2011). Thus, across the relatively small geographic area of the Israel landscape there is a wide gradient of social structures, which is particularly apparent between the inland populations and those in the coastal plain. These differences in social structure may explain the genetic differentiation between lineages, as they may create reproductive barriers between social forms. Recently, an association between social structure and genetic background was reported in two different ant species, the fire ant *Solenopsis invicta* and the Alpine silver ant *Formica selysi* (Wang et al., 2013; Purcell et al., 2014). In those species, the two distinct forms of social structures (i.e., monogynous vs. polygynous) are determined by the two haplotypes of a genomic region linked in a supergene, termed the "social chromosome". In the present study, microsatellite analysis did not reveal any evidence of reproductive isolation between social forms suggesting that it is behavioral flexibility rather than a genetic basis that underlies the variation in social phenotype. However, we cannot completely ruled out the hypothesis that differential reproductive strategies under distinct ecological factors may be a driving force behind speciation process, despite the absence of global reduction of gene flow (Nosil and Schluter, 2011). Under such speciation-with-gene-flow scenario, differential reproductive strategies may yield genetic divergences in genomic regions under divergent selection, while the rest of the neutral or weakly selected genome is homogenized by gene flow (Feder et al., 2012; Abbott et al., 2013). Whether ecological factors, social structure or nest architecture (i.e., monodomous to supercolonies) are genetically influenced, and whether potential outlier genomic regions indicate some locally-adapted speciation are clearly worth additional future study of the *Cataglyphis bicolor* group.

This difference in mating system may also have profound implications in dispersal extents and hence in the variation of natural genetic diversity along the Israeli landscape. In social insects, new colonies may be established in two different ways (Cronin et al., 2013). During independent colony foundation, the new queens proceed in long-range dispersal through nuptial flights. This strategy is usually associated to monogyne colonies and lead to high genetic mixing in the population (Ross and Shoemaker, 1997). On the other hand, during dependent foundation, young queens mate close to or in the natal nest and disperse by foot with a worker force to establish a new colony nearby. This short-range dispersal results in a pattern of genetic isolation-by-distance within populations and enhances genetic differentiation between populations (Liautard and Keller, 2001; Clemencet et al., 2005; Leppänen et al., 2013). In our case, *C. savignyi* mtDNA haplotypes show weak divergence (0.50%) for the large sampling area, contrasting with *C. niger* that exhibit higher divergence (1.39%) despite its reduced sampling range (more than 6 times smaller than the one of *C. savignyi*). This difference may stem from distinct mode of colonies founding between the polygynous *C. niger* and the monogynous *C. savignyi* (Leniaud et al., 2011).

Finally, a comparison with *mt*DNA sequences from GenBank revealed that *C. savignyi/C. niger* species are restricted to Israel and Egypt, and strongly differ from the homonymous *C. savignyi* species in Mali and Tunisia (Knaden et al., 2012).

4.2. POL and TSI mtDNA haplotypes

Our results also revealed two highly divergent mitochondrial haplotypes, one located in the central coastal plain (i.e. POL & NET) and the other in lower Galilee (i.e. TSI). These haplotypes were found in very restricted areas of one or two neighboring localities, and do not belong to any of the groups studied to date in Israel. Remarkably, these highly *mt*DNA-distinct individuals do not carry diagnostic nuclear genes or possess distinct microsatellite markers, but are rather similar to the surrounding individuals, suggesting a constant gene flow between them and the surrounding populations. We tend to exclude the possibility of misamplification of nuclear copies of mitochondrial genes (NUMTs) rather than being true mitochondrial ones (Martins et al., 2007), because this pattern was found on both the COI and CytB markers belonging to the two eponymous genes that are highly distant on the mitochondrial genome, rendering it unlikely that these distant genes had experienced the same duplication. Furthermore, no stop codons were found in the distant haplotypes, suggesting that they belong to the true mitochondrial genome (D'Errico et al., 2004). We hypothesize that mitochondrial introgression occurred, derived from the invasion by an interspecific queen with a highly divergent mitochondrial haplotype. The unequal abundance of parental species facilitates potential interspecific mating with males from surrounding populations (Gardner, 1997; Vollmer and Palumbi, 2002). Across generations, the high levels of mixing between invader queens and local males may swamp the nuclear gene pool of the invader and lead to two species merging into one, while the highly divergent *mt*DNA remains maternally and asexually preserved (Seehausen, 2006; Excoffier et al., 2009; Coleman et al., 2014).

4.3. Conclusion

The present phylogenetic study reframes our understanding of the recent history of *Cataglyphis* desert ants of the *bicolor* group in Israel, highlighting how highly diverse habitats in the Israeli landscape, and differences in mating systems and social structures, may shape speciation processes in this group. The recent diversification of the *C. savignyi*, *C. niger*, and *C. drusus* lineages associated with profound differences in social structures and dispersal abilities offers an excellent framework from which to study the evolutionary processes underlying speciation. Does social polymorphism create mechanisms of reproductive barriers between social forms that potentially lead to the separation of these lineages; or does this geographical subgrouping result from differences in dispersal aptitudes between polygynous and monogynous populations? These evolutionary questions are expected to encourage further research into obtaining genomic datasets and biological data on social systems for a broader range of *Cataglyphis* species of this group. The consequent studies of social structure evolution will bring us even closer to an understanding of why the ant genus *Cataglyphis* exhibits such a diverse array of reproductive strategies.

Acknowledgments

We are grateful to K. Kiran and A. Ionescu-Hirsch for specimen sampling, and to S. Aron, M. Pearcy, L. Leniaud, T. Simon, and M. Orlova for their help in genetic and chemical analyses. Our work was supported by a grant from the Israel Science Foundation (ISF grant # 844/13 to AH), a fellowship from the George S. Wise Faculty of Life Sciences and David Furth's grant for Systematic Entomology (to PAE). We are grateful for the support received from The Steinhardt Museum of Natural History, Israel National Center for Biodiversity Studies and Department of Zoology, Tel Aviv. We thank Naomi Paz for editorial assistance.

Data accessibility: DNA sequences: GenBank accession numbers MF072038 to MF072315.

Authors contributions

PAE, RZ and AH designed the research; field collection was carried out by PAE, RZ and AH; PAE, RZ and AH performed the research; PAE analyzed the data; PAE and AH wrote the paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.07.024.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., et al., 2013. Hybridization and speciation. J. Evol. Biol. 26, 229–246.

Agosti, D., 1990. Review and reclassification of *Cataglyphis* (Hymenoptera, Formicidae). J. Nat. Hist. 24, 1457–1505.

Aron, S., Mardulyn, P., Leniaud, L., 2016a. Evolution of reproductive traits in *Cataglyphis* desert ants: mating frequency, queen number, and thelytoky. Behav. Ecol. Sociobiol. 1–13.

Aron, S., Lybaert, P., Baudoux, C., Vandervelden, M., Fournier, D., 2016b. Sperm production characteristics vary with level of sperm competition in *Cataglyphis* desert ants. Funct. Ecol. 30, 614–624.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., et al., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148–155.

Boulay, R., Aron, S., Cerdá, X., Doums, C., Graham, P., Hefetz, A., Monnin, T., 2017. Social life in arid environments: the case study of *Cataglyphis* ants. Annu. Rev. Entomol. 62, 305–321.

Brady, S.G., Schultz, T.R., Fisher, B.L., Ward, P.S., 2006. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc. Natl. Acad. Sci. USA 103, 18172–18177.

Carstens, B.C., Knowles, L.L., 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* Grasshoppers. Syst. Biol. 56, 400–411.

Clemencet, J., Viginier, B., Doums, C., 2005. Hierarchical analysis of population genetic structure in the monogynous ant *Cataglyphis* cursor using microsatellite and mitochondrial DNA markers. Mol. Ecol. 14, 3735–3744.

Coleman, R.R., Gaither, M.R., Kimokeo, B., Stanton, F.G., Bowen, B.W., et al., 2014. Large-scale introduction of the Indo-Pacific damselfish *Abudefduf vaigiensis* into Hawai'i promotes genetic swamping of the endemic congener *A. abdominalis*. Mol. Ecol. 23, 5552–5565.

Cronin, A.L., Molet, M., Doums, C., Monnin, T., Peeters, C., 2013. Recurrent evolution of dependent colony foundation across eusocial insects. Annu. Rev. Entomol. 58, 37–55.

Dahbi, A., Lenoir, A., Tinaut, A., Taghizadeh, T., Francke, W., Hefetz, A., 1996. Chemistry of the postpharyngeal gland secretion and its implication for the phylogeny of Iberian *Cataglyphis* species (Hymenoptera: Formicidae). Chemoecology 7, 163–171.

Dahbi, A., Hefetz, A., Lenoir, A., 2008. Chemotaxonomy of some *Cataglyphis* ants from Morocco and Burkina Faso. Biochem. Syst. Ecol. 36, 564–572.

Darras, H., Leniaud, L., Aron, S., 2014. Large-scale distribution of hybridogenetic lineages in a Spanish desert ant. Proc. R Soc. B Biol. Sci. 281.

Dellicour, S., Mardulyn, P., 2014. Spads 1.0: a toolbox to perform spatial analyses on DNA sequence data sets. Mol. Ecol. Resour. 14, 647–651.

Dellicour, S., Flot, J.-F., 2015. Delimiting species-poor datasets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. Syst. Biol. 64, 900–908.

D'Errico, I., Gadaleta, G., Saccone, C., 2004. Pseudogenes in metazoa: origin and features. Brief. Funct. Genomic Proteomic. 3, 157–167.

De Queiroz, K., 2007. Species Concepts and Species Delimitation. Syst. Biol. 56, 879–886.

Doyle, J.J., 1995. The irrelevance of allele tree topologies for species delimitation, and a non-topological alternative. Syst. Bot. 20, 574–588.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Dupanloup, I., Schneider, S., Excoffier, L., 2002. A simulated annealing approach to define the genetic structure of populations. Mol. Ecol. 11, 2571–2581.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113–116.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver.3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567.

Excoffier, L., Foll, M., Petit, R.J., 2009. Genetic consequences of range expansions. Annu Rev. Ecol. Evol. Syst. 40, 481–501.

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28, 342–350.

Flot, J.-F., 2007. Champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. Mol. Ecol. Notes 7, 974–977.

Flot, J.-F., 2010. Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol. Ecol. Resour. 10, 162–166.

Flot, J.-F., Couloux, A., Tillier, S., 2010. Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "field for recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. BMC Evol. Biol. 10, 372.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27, 480–488.

Fujisawa, T., Barraclough, T.G., 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Syst. Biol. 62, 707–724.

Gardner, J.P.A., 1997. Hybridization in the sea. Adv. Mar. Biol. 31, 1–78.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Holder, M.T., Anderson, J.A., Holloway, A.K., 2001. Difficulties in detecting hybridization. Syst. Biol. 50, 978–982.

Holland, M.M., Parson, W., 2011. GeneMarker® HID: a reliable software tool for the analysis of forensic STR data. J. Forensic Sci. 56, 29–35.

Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V., Huber, K.T., 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evol. Biol. 8, 202.

Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. Evolution 56, 1557–1565.

Ionescu, A., Eyer, P.-A., 2016. Notes on *Cataglyphis* Foerster, 1850 species belonging to the *bicolor* species-group in Israel; and a description of a new species. Israel J. Entomol. 46, 109–131.

Joly, S., McLenachan, P., Xa, A., Lockhart, P., et al., 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174, E54–E70.

Keegans, S.J., Morgan, E.D., Agosti, D., Wehner, R., 1992. What do glands tell us about species? A chemical case study of *Cataglyphis* ants. Biochem. Syst. Ecol. 20, 559–572.

Knaden, M., Tinaut, A., Stökl, J., Cerdá, X., Wehner, R., 2012. Molecular phylogeny of the desert ant genus *Cataglyphis* (Hymenoptera: Formicidae). Myrmecol. News 16, 123–132.

Knaden, M., Tinaut, A., Cerda, X., Wehner, S., Wehner, R., 2005. Phylogeny of three parapatric species of desert ants, *Cataglyphis bicolor*, *C. viatica*, and *C. savignyi*: a comparison of mitochondrial DNA, nuclear DNA, and morphological data. Zoology 108, 169–177.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G., Savolainen, V., 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. USA 105, 2923–2928.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. J. Statistic Soft 25, 1–18.

Leniaud, L., Heftez, A., Grumiau, L., Aron, S., 2011. Multiple mating and supercoloniality in *Cataglyphis* desert ants. Biol. J. Linnean Soc. 104, 866–876.

Lenoir, A., Aron, S., Cerda, X., Hefetz, A., 2009. *Cataglyphis* desert ants: a good model for evolutionary biology in Darwin's anniversary year—a review. Israel J. Entomol.

Leppänen, J., Vepsäläinen, K., Anthoni, H., Savolainen, R., 2013. Comparative phylogeography of the ants *Myrmica ruginodis* and *Myrmica rubra*. J. Biogeo. 40, 479–491.

Liautard, C., Keller, L., 2001. Restricted effective queen dispersal at a microgeographic scale in polygynous populations of the ant *Formica exsecta*. Evolution 55, 2484–2492.

Lumbsch, H.T., Leavitt, S.D., 2011. Goodbye morphology? A paradigm shift in the delimitation of species in lichenized fungi. Fungal Divers. 50, 59–72.

Mantel, N., 1967. The detection of disease clustering and generalized regression approach. Cancer Resour. 27, 209–220.

Martins, J., Solomon, S.E., Mikheyev, A.S., Mueller, U.G., Ortiz, A., et al., 2007. Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). Insect Mol. Biol. 16, 777–784.

Mayr, E., 1942. Systematics and the Origin of Species. Columbia University Press, New York.

Miller, M.P., 2005. Alleles In Space (AIS): Computer Software for the Joint Analysis of Interindividual Spatial and Genetic Information. J. Hered. 96, 722–724.

Nosil, P., Schluter, D., 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26, 160–167.

Oldham, N.J., Morgan, E.D., Agosti, D., Wehner, R., 1999. Species recognition from postpharyngeal gland contents of ants of the *Cataglyphis bicolor* group. J. Chem. Ecol. 25, 1383–1393.

Papadopoulou, A., Anastasiou, I., Vogler, A.P., 2010. Revisiting the insect mitochondrial molecular clock: the Mid-Aegean Trench calibration. Mol. Biol. Evol. 27, 1659–1672.

Pearcy, M., Clemencet, J., Chameron, S., Aron, S., Doums, C., 2004. Characterization of nuclear DNA microsatellite markers in the ant Cataglyphis cursor. Mol. Ecol. Notes 4, 642–644.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst. Biol. 55, 595–609.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol. Ecol. 21, 1864–1877.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., Chapuisat, M., 2014. Convergent genetic architecture underlies social organization in ants. Curr. Biol. 24, 2728–2732.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Radchenko, A., 2001. The phylogeny and faunogenesis of the genus *Cataglyphis* Foerster (Hymenoptera, Formicidae). Entomol. Obozrenie. 80, 885–895.

Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164, 1645–1656.

Reid, N., Carstens, B., 2012. Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent

model. BMC Evol. Biol. 12, 196.
Ross, K.G., Shoemaker, D.D., 1997. Nuclear and mitochondrial genetic structure in two social forms of the fire ant *Solenopsis invicta*: insights into transitions to an alternate social organization. Heredity 78, 590–602.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Saar, M., Leniaud, L., Aron, S., Hefetz, A., 2014. At the brink of supercoloniality: genetic, behavioral and chemical assessments of population structure of the desert ant *Cataglyphis niger*. Front. Ecol. Evol. 2.
Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, A., 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55, 421–438.
Seehausen, O., 2006. Conservation: losing biodiversity by reverse speciation. Curr. Biol. 16, R334–R337.
Stephens, M., Smith, N., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.
Ströher, P.R., Li, C., Pie, M.R., 2013. Exon-primed intron-crossing (EPIC) markers as a tool for ant phylogeography. Rev. Brasil. Entomol. 57, 427–430.
Sullivan, J.P., Lavaqué, S., Hopkins, C.D., 2002. Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). Evolution 56, 597–616.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., et al., 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31, 21–32.
Vollmer, S.V., Palumbi, S.R., 2002. Hybridization and the Evolution of Reef Coral Diversity. Science 296, 2023–2025.
Vonshak, M., Ionescu-Hirsch, A., 2009. A checklist of the ants of Israel (Hymenoptera: Formicidae). Israel J. Entomol. 39, 33–55.
Walsh, P.S., Metzger, D.A., Higuchi, R., 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10, 506–513.
Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., et al., 2013. A Y-like social chromosome causes alternative colony organization in fire ants. Nature 493, 664–668.
Wiens, J.J., 2007. Species delimitation: new approaches for discovering diversity. Syst. Biol. 56, 875–878.
Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl Acad. Sci. USA 107, 9264–9269.
Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869–2876.