Zoologica Scripta

WILEY | KUNGL. VETENSKAPS-AKADEMIEN

ORIGINAL ARTICLE OPEN ACCESS

Evolution of *Crematogaster sordidula* (Hymenoptera, Formicidae) Ants in the Mediterranean Region During Plio-Pleistocene Climatic Changes

Jody H. Voges[1,2] | Sebastian Salata[3] | Lech Borowiec[3] | Jelena Bujan[4] | Christos Georgiadis[5] | Celal Karaman[6] | Kadri Kiran[6] | Albena Lapeva-Gjonova[7] | Claude Lebas[8] | Mattia Menchetti[1,9] | Enrico Schifani[9,10] | Ahmed Taheri[11] | Bonnie B. Blaimer[1]

[1]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Center for Integrative Biodiversity Discovery, Berlin, Germany | [2]Faculty of Life Sciences, Department of Biology, Humboldt Universität Zu Berlin, Berlin, Germany | [3]Myrmecological Laboratory, Department of Biodiversity and Evolutionary Taxonomy, University of Wroclaw, Wroclaw, Poland | [4]Division for Marine and Environmental Research, Ruđer Bošković Institute, Zagreb, Croatia | [5]Section of Zoology-Marine Biology & Zoology Museum, Department of Biology, National and Kapodistrian University of Athens, Athens, Greece | [6]Faculty of Science, Department of Biology, Trakya University, Edirne, Türkiye | [7]Department of Zoology and Anthropology, Sofia University, Sofia, Bulgaria | [8]Impasse del Ribas, Canohes, France | [9]Institut de Biologia Evolutiva (CSIC-Univ. Pompeu Fabra), Barcelona, Spain | [10]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy | [11]LESCB, URL-CNRST No. 18, Faculty of Sciences, Abdelmalek Essaadi University, Tetouan, Morocco

Correspondence: Jody H. Voges (vogesjody@gmail.com)

Received: 1 August 2025 | **Revised:** 12 February 2026 | **Accepted:** 13 February 2026

ABSTRACT

Understanding insect responses to global climate change involves identifying strategies used during past climate oscillations. Phylogeography offers a powerful framework to unravel how historical climatic and geological events have shaped the spatial genetic patterns of species, providing critical insights into evolutionary processes, whereas population genomic approaches enable finer-scale resolution of demographic history and genetic connectivity. We focus on an abundant ant species, *Crematogaster sordidula*, endemic to the Mediterranean region, a biodiversity hotspot experiencing faster warming than the global average. Using a phylogenomic dataset of ultraconserved elements (UCEs) from 123 individuals, we reconstructed population-level relationships with maximum-likelihood and coalescent methods and estimated a time-calibrated phylogeny to investigate the timing of major divergence events. We inferred population structure by generating mitochondrial haplotype networks and analysing admixture and regional genetic diversity and differentiation on the basis of SNPs obtained from UCE data. We recovered four major lineages that diverged during the Pliocene: a Western Mediterranean clade, an Eastern Aegean region clade, an Anatolia-Türkiye clade, and a Central-Eastern Mediterranean clade, with the latter three together forming a larger Eastern Mediterranean clade. This population structure confirms the east–west distribution pattern found for many taxa in the Mediterranean. Genetic diversity was highest in southern regions, supporting the hypothesis that these areas acted as refugia during Pleistocene climatic oscillations. Within the Eastern Mediterranean clade, population structure and gene flow indicate complex demographic histories possibly shaped by the region's dynamic palaeogeography and paleoclimatic history. Together, these results provide a foundation for understanding the species' potential responses to ongoing environmental changes.

1 | Introduction

The Mediterranean region is one of the world's biodiversity hotspots (Myers et al. 2000) with a high diversity of species and remarkable levels of endemism, reaching 50% in plants and more than 60% in some animal groups such as fishes and amphibians (Blondel et al. 2010). The region lies at the intersection of two major landmasses, Eurasia and Africa, and hosts a large topographical diversity, from small islands and islets starting at sea level up to continental mountain ranges with altitudes over 4000 m. The Mediterranean Basin borders 24 countries and stretches over ca. 3800 km west–east and 1000 km north–south with a total area of 2.5 million km^2, harbouring 11,879 islands and islets with a combined coastline of 24,600 km (Blondel et al. 2010). The area is divided into two subregions by the Strait of Sicily, the western and the eastern Mediterranean Basin (Cartes et al. 2004). The current configuration of the Mediterranean Sea was established after the opening of the Strait of Gibraltar at the beginning of the Pliocene, around 5.3 Ma (Periáñez and Abril 2015).

Paleoclimate changes and geological events throughout the Pliocene and Pleistocene have profoundly shaped the habitat heterogeneity and species diversity of the Mediterranean region (Blondel et al. 2010). The present-day Mediterranean climate is characterised by hot, dry summers and mild, humid winters (Lionello et al. 2006), but only became established throughout the region around 2.8 Ma and even then was still subject to frequent climatic oscillations during glacial and interglacial periods (Poulakakis et al. 2015). Although the Mediterranean area was not covered by ice sheets during glaciations throughout the Pliocene and Pleistocene, it was highly impacted by their distribution in northern latitudes (Gat and Magaritz 1980). When glaciers covered the higher latitudes of the northern hemisphere and major parts of northern Eurasia, parts of the Mediterranean, such as the Iberian and Balkan Peninsulas, as well as Sicily and Crete, acted as southern refugia for a variety of taxa during that time (Hewitt 2001; Médail and Diadema 2009). Formation of glaciers in the north also led to a decreased sea level in the Mediterranean area, re-connecting previously isolated islands with the Mediterranean mainland and allowing for biotic exchange (Benítez-Benítez et al. 2021). Conversely, during interglacials, the increase of the eustatic sea level in the Mediterranean basin caused the separation and fragmentation of islands and islets (Dermitzakis 1990).

Given its complex and interesting paleoclimatic history, the Mediterranean region is an ideal setting to study the evolution and biogeography of animals in the context of climatic changes. A large body of research exists on the phylogeography of vertebrates such as reptiles and amphibians (e.g., Jablonski et al. 2016; Kornilios et al. 2019; Salvi et al. 2013; Sherpa et al. 2024), but invertebrates have received—relative to their diversity—much less attention (Poulakakis et al. 2015). For insects, the Mediterranean region is the most diverse area within Europe and is inhabited by 75% of all described European species (Blondel et al. 2010). Previous work on the evolution and phylogeography of the Mediterranean insect fauna has focused mainly on Orthoptera (Allegrucci et al. 2009, 2011; Borissov et al. 2020, 2021, 2023; Hochkirch et al. 2023), Coleoptera (Avtzis et al. 2021; Fattorini 2002; Mas-Peinado et al. 2018; Papadopoulou et al. 2009; Trichas et al. 2020) and Lepidoptera (mainly butterflies, e.g., Cassar et al. 2025; Dapporto et al. 2022, 2024; Habel et al. 2005, 2008; Kühne et al. 2017), whereas the phylogenetic histories of other groups such as the species-rich Hymenoptera remain less studied.

Ants (Hymenoptera: Formicidae) are one of the most abundant and diverse groups of insects (Hölldobler and Wilson 1990) and are an ideal subject to study evolutionary responses of species to environmental changes in the Mediterranean region because of their diversity and high levels of endemism (Tinaut and Ruano 2021). A checklist for the ants of Europe, the Mediterranean Basin and adjacent countries tallies 1261 species (Borowiec 2014, 2021), of which a large part are endemic to the Mediterranean. Indeed, the Mediterranean region is a global hotspot for ant diversity, especially concerning rare endemic species (Kass et al. 2022). Ant diversity in the region has been deeply influenced by geographic and climatic constraints that evolved along with its complex history (Wang et al. 2023). Over the last decade, molecular data have increasingly contributed to shedding new light on the diversity and biogeography of Mediterranean ants, whether examining patterns within particular species (Csősz et al. 2025), groups of species (Csősz et al. 2015), genera (Schifani et al. 2022; de Lecocq Pletincx et al. 2024), or geographic areas (Blatrix et al. 2020; Schär et al. 2020). Another significant step forward in understanding the evolutionary history of European and Mediterranean ants is presented by a comprehensive study estimating gene trees and genetic maps on the basis of the mitochondrial gene *cytochrome oxidase I* (COI) for 480 European ant species, many of them endemic to the Mediterranean Basin, providing intriguing further insights into the phylogeography of ants in this region (Menchetti et al. n. d.)

Crematogaster sordidula (Nylander 1849; Figure 1), in some regions also known as the Sordid cocktail ant, is a widespread ant species and the sole described member of the subgenus *Orthocrema* within the diverse genus *Crematogaster* occurring in the Mediterranean region. The species has been recorded from large parts of the southern Palearctic region, spanning Southern Europe and Northern Africa to Southern Russia, Kazakhstan, and Iran (Guénard et al. 2017), and is particularly thriving in Mediterranean habitats from sea level up to medium altitudes. Workers of this species are relatively small, typically less than 3 mm in length, and colonies in the Mediterranean region predominantly inhabit open, exposed habitats, such as grasslands, macchia, and forest edges (Seifert 2018). The species prefers warm microhabitats, nesting in the soil under stones or beneath plants. Its widespread distribution in diverse environments at different elevations and across various Mediterranean ecosystems demonstrates its ecological flexibility, making it an ideal candidate for studying species distribution patterns and the evolutionary history of ants in the region. First genetic results for *C. sordidula* on the basis of *cytochrome oxidase I* (COI) showed a deep divergence between western and eastern Mediterranean populations and hinted at an interesting population structure, particularly in the eastern Mediterranean region (Menchetti et al. n. d.)

In this study, we generate the first comprehensive population-level phylogeny for Mediterranean ants. We utilise the

FIGURE 1 | *Crematogaster sordidula* (Nylander, 1849) workers. (A) A *C. sordidula* ground nest under a stone, Albania. Photo credit: Bonnie B. Blaimer; (B) close-up of *C. sordidula* workers on the ground, France. Photo credit: Benjamin Zelvelder.

widespread ant species *C. sordidula* and a dataset comprising more than 2000 nuclear genomic markers (ultraconserved elements or UCEs) to conduct a dating and biogeographic analysis to investigate lineage diversification. We further use SNPs to investigate the structure, genetic diversity, and connectivity among populations, and additionally explore haplotype diversity for a subset of samples on the basis of mitochondrial data to gain a comprehensive picture of the species' evolutionary history. We assess the presence of ancient divergence associated with geographical distance between the eastern and western Mediterranean Basin, as previously recovered for *C. sordidula* and many other ant taxa on the basis of mitochondrial data. We test for gene flow between populations and investigate whether lineage diversification patterns are consistent with the isolation of populations in possible refugial areas, such as Sicily, Crete, and the Iberian, Italian, and Balkan Peninsulas, during episodes of glaciation. We discuss our results in relation to these predictions and in the light of the complex paleogeographic history of the Mediterranean region and the influence of past climate changes.

2 | Materials and Methods

2.1 | Taxon Sampling

Our dataset comprised 123 ingroup samples of *C. sordidula* and three outgroup taxa from the same genus: *Crematogaster* (*Orthocrema) longipilosa* Forel, 1907, and *C.* (*Orthocrema*) *nigropilosa* Mayr, 1870, as more closely related members of the same subgenus from other biogeographic regions, and *C.* (*Crematogaster*) *scutellaris* (Olivier, 1791) as the outer outgroup. *Crematogaster sordidula* exhibits significant intraspecific morphological variation across its distribution range, which has resulted in the description of several subspecies. These include the currently valid taxa *C. sordidula aeolia* Forel, 1911, and *C. sordidula osmanica* Santschi, 1921, described from Türkiye, *C. sordidula marocana* Santschi, 1921, from Morocco, and *C. sordidula molongori* Weber, 1943, from Sudan. The taxonomic validity of these subspecies is unclear, and distinctions of subspecies from the nominal species are difficult. For this study, we therefore chose to focus on the nominal species only, although some of our samples from Türkiye had been tentatively identified as *C. sordidula aeolia*. The taxonomic status of *C. sordidula* and its subspecies will be resolved in the future in a separate study.

Table S1 contains voucher information for all 126 samples. We included samples from 10 countries (Albania, Bulgaria, Croatia, France, Greece, Italy, Montenegro, Morocco, Spain, and Türkiye) distributed over an area of around 3 million square meters. Specimens were sampled from sea level to a maximum elevation of about 1700 m (Table S1). We were not able to include specimens from the Mediterranean countries Algeria, Israel, Lebanon, Syria, and Tunisia in our study. All specimens included in this study were collected in accordance with local regulations, and all necessary permits were obtained.

2.2 | Molecular Data Collection

We extracted DNA from specimens using a Qiagen DNeasy Blood & Tissue kit (Qiagen, Valencia, CA). A slightly modified version of the manufacturer's protocol was followed. We either extracted DNA destructively by pulverising the entire ant (mostly used for pupae and in cases where multiple individuals from the same colony series were available) or non-destructively (mostly used for adults), enabling us to re-mount voucher specimens after extractions. For all destructively extracted specimens, we pinned and stored a specimen from the same colony sample as a voucher for future morphological consideration. Vouchers are stored at the MfN Berlin, or the University of Wroçlaw. The DNA quality was checked first by measuring DNA concentration with a Qubit fluorometer (High sensitivity kit, Life Technologies Inc., Carlsbad, CA), and second by running 5 μL of DNA eluate on a 1%-Agarose gel for 1 h at 120 V.

We used a Qsonica sonicator (Q800R3; Qsonica Inc., Newton, CT) to shear < 5–100 ng (64.6 ng mean) DNA to a target size of

approximately 500–600 bp from 0 to 90 s (amp = 25, pulse = 10), depending on prior DNA degradation. This sheared DNA was used as input for a modified genomic DNA library preparation protocol using a Kapa Hyper Prep Library Kit (Kapa Biosystems Inc., Wilmington, MA) and incorporating bead-based cleanup steps (Fisher et al. 2011), a generic SPRI substitute (Rohland and Reich 2012), and TruSeq-style adapters during adapter ligation (Faircloth and Glenn 2012), following Faircloth et al. (2012). After 13 PCR cycles, we obtained post-library preparation DNA concentrations between 1.04 to 42.7 ng/μL (mean 15.89 ng/μL), quantified using a Qubit fluorimeter (Broad range kit; Life Technologies Inc.). We pooled libraries at equimolar concentrations containing eight to 10 samples prior to target enrichment, aiming at total target DNA input of 500 ng.

We enriched these pools using a published Hymenoptera bait set (myBaits UCE Hymenoptera 2.5Kv2P; Branstetter et al. 2017), which targets 2590 UCE loci shared among Hymenoptera and contains 31,829 unique baits. We followed an enrichment protocol by Blumenstiel et al. (2010), modified by Faircloth et al. (2012) and ArborBiosciences (MYbaits version 4.01). Finally, we performed qPCR quantification of enriched library pools to obtain accurate DNA concentrations of the pools using a SYBR FAST qPCR kit (Kapa Biosystems Inc., Wilmington, MA) and a CFX Opus Real-Time PCR System (Bio-Rad Laboratories Inc.). Sequencing was performed on an Illumina Novaseq 6000 instrument (Novogene, Cambridge, UK).

2.3 | Bioinformatic Processing and Alignment of UCE Loci

Sequence data processing was performed using PHYLUCE v1.7.2 (Faircloth 2015), a pipeline designed for UCE data, as outlined in brief in the following. All PHYLUCE settings were left at default values unless mentioned otherwise. The demultiplexed raw sequence reads were trimmed using ILLUMIPROCESSOR (Faircloth 2013), which incorporates TRIMMOMATIC (Bolger et al. 2014). The adapter-trimmed reads were used for de novo assembly with SPADES v3.15.4 (Bankevich et al. 2012) using the phyluce_assembly_assemblo_spades.py script. Contigs representing UCE loci were identified using phyluce_assembly_match_contigs_to_probes.py (min. coverage = 50, min. identity = 80). Sequence coverage statistics were calculated for UCE contigs using phyluce_assembly_get_fastq_lengths (Table S2). The data were aligned with MAFFT within the PHYLUCE environment using phyluce_assembly_sequap_align.py (no-trim, min-length = 100). To trim poorly aligned internal and flanking regions, Gblocks (Talavera and Castresana 2007) was used by running phyluce_assembly_get_gblocks_trimmed_from_untrimmed.py, with reduced stringency parameters (b1:0.5, b2:0.5, b3:12, b4:7). Parameter settings follow the description by Blaimer et al. (2015). We created two different datasets selecting subsets of UCE alignments depending on the amount of captured UCE loci per taxon: (a) 80% complete (loci with > 102 of 126 taxa represented) and (b) 90% complete (loci with > 115 of 126 taxa represented). All scripts, default settings, and a more detailed description of the steps of the PHYLUCE pipeline can be found at https://phyluce.readthedocs.io/en/latest/tutorials/.

For both datasets, we produced concatenated alignments and applied different trimming and partitioning strategies. First, we identified and removed poorly aligned sequences using three distance cut-off values (0.95, 0.97, 0.98) in Spruceup v2020.2.19 (M. L. Borowiec 2019), a Python package to discover and remove poorly aligned sequences within alignments. Second, we partitioned the data matrices using the Sliding-Window Site Characteristics Entropy (SWSC-EN) algorithm, which uses site characteristics such as entropy to generate partitions that account for heterogeneity between rates of core and flanking regions of UCEs (Tagliacollo and Lanfear 2018). In the next step, PartitionFinder2 v.2.1.1 (Lanfear et al. 2016) was used to select best-fit partitioning schemes and models of evolution for phylogenetic analyses.

We thus used nine different combinations of spruceup trimming and data partitioning to perform concatenated maximum likelihood analyses: (i) 80% completeness matrix, no partitioning (80%-untrimmed_np); (ii) 80% completeness matrix, partitioned (80%-untrimmed_p); (iii) 80% completeness matrix, partitioned, spruceup trimming with 0.95 cut-off (80%-0.95-spruceup); (iv) 80% completeness matrix, partitioned, spruceup trimming with 0.97 cut-off (80%-0.97-spruceup); (v) 80% completeness matrix, partitioned, spruceup trimming with 0.98 cut-off (80%-0.98-spruceup); (vi) 90% completeness matrix, no partitioning (90%-untrimmed); (vii) 90% completeness matrix, partitioned, spruceup trimming with 0.95 cut-off (90%-0.95-spruceup); (viii) 90% completeness matrix, partitioned, spruceup trimming with 0.97 cut-off (90%-0.97-spruceup); (ix) 90% completeness matrix, partitioned, spruceup trimming with 0.98 cut-off (90%-0.98-spruceup). All alignment characteristics, including length, missing data percentages, variable sites, parsimony informative sites, and AT/GC content, are listed in Table S3.

2.4 | Phylogenetic Inference From UCE Loci

We inferred phylogenetic relationships within *C. sordidula* on the basis of the above-described alignments with the likelihood-based program IQ-TREE v.1.6.8 (Nguyen et al. 2014). Analyses were performed using maximum likelihood (ML) best-tree and ultrafast bootstrap searches ($N = 1000$), employing model selection for unpartitioned matrices using the ModelFinder option, while implementing the GTR + G model for data subsets in partitioned matrices. The alrt option was additionally used to assess branch supports with single-branch tests, an implementation of the SH-like approximate likelihood ratio test (Guindon et al. 2010), run with several bootstrap replicates for SH-aLRT = 1000. Analyses were rooted using the most distantly related outgroup taxon *C. scutellaris* (Blaimer 2012). All resulting phylogenetic trees are available in the Figures S1–S5.

To perform coalescent analysis, gene trees were created from spruceup trimmed (0.98 cut-off) locus alignments with 90% of

taxa present (=1921 loci), using the inferring gene/locus trees workflow in IQ-TREE. After running the IQ-TREE locus analyses, the gene.tree file containing all the input trees in NEWICK format was then used to perform coalescent species-tree analyses in ASTRAL-III v.5.7.8 (Zhang et al. 2018), using default parameters.

2.5 | SNP Calling and Alignment for Phylogenomic Analyses

Allelic sequences are useful for phylogenetic studies to improve the estimation of gene trees, species trees, and divergence times (Andermann et al. 2018; Garrick 2010; Lischer et al. 2014; Potts et al. 2014). We used the workflow from Andermann et al. (2018), implemented in PHYLUCE for allele phasing from UCE data. This workflow maps reads using bwa (Li and Durbin 2009), and sorts and phases reads using SAMtools v0.1.19 (Li et al. 2009) on the basis of a dynamic programming algorithm that uses read connectivity across multiple variable sites to determine the two phases of any given diploid locus (He et al. 2010). The output of the allele phasing workflow is a consensus sequence for each of the sorted BAM files, the allelic sequences. In the next step, the allele sequences were aligned using phyluce_align_seqcap_align in PHYLUCE. We inferred phylogenetic relationships for the SNP alignment using IQ-TREE v.1.6.8 (Nguyen et al. 2014) to confirm the tree topology estimated from concatenated alignments. Analyses were performed using a maximum likelihood (ML) best-tree and ultrafast bootstrap search ($N = 1000$), employing model selection using the ModelFinder option. Analyses were rooted using the most distantly related outgroup taxon *C. scutellaris* (Blaimer 2012).

After confirming the previously inferred topology from unphased data with the results from the phased data, we produced a condensed dataset with only two selected taxa per lineage to reduce computational time for divergence dating analyses. We decided on 14 discrete units on the basis of phylogenetic divergence in our dataset and assigned all samples to these. Samples were sorted by the number of reads, and for each lineage, two samples with the highest data amounts were chosen. For the Moroccan lineage, only three samples were available, with two of them having a low number of reads. Only one sample from this lineage was included in the SNP alignment to reduce the number of missing sites in the final matrix. Information on the samples and lineages is listed in Table S4. The following steps were performed on the reduced dataset, comprising 27 samples with two allele sequences per sample, resulting in a matrix with 54 sequences.

We used SNP-SITES (Page et al. 2016) and followed the pipeline from Brown et al. (https://github.com/jasonleebrown/UCE_phyluce_pipeline), while modifying this to run on a UNIX system. This pipeline takes phased UCE data as input and first calls all SNPs per locus, then counts and randomly selects one SNP per locus, and concatenates these SNPs into one alignment. We reduced the heterozygous information to a single variant SNP per locus, allowing the use of Bayesian Multispecies Coalescent methods. The final SNP alignment consisted of exactly one variable position per locus with two states to ensure unlinked SNPs, excluding positions with missing data or ambiguities. This produced a SNP dataset of 1395 unlinked SNPs. We ran an IQ-TREE analysis using a maximum likelihood (ML) best-tree and ultrafast bootstrap search ($N = 1000$) on the basis of this SNP alignment to obtain a population-level tree as input for divergence dating.

2.6 | Divergence Dating

Dating a phylogeny with both inter- and infraspecific samples requires using the same prior for all branching processes and can lead to an over- or underestimation of divergence times (Angelis and Dos Reis 2015). We were primarily interested in dating intraspecific divergences within our ingroup taxa to infer the biogeographical history of *C. sordidula* on a population level; however, this required a calibration on the crown-group node for *C. sordidula*. In the absence of suitable fossils or previous molecular dating estimates for *C. sordidula*, we decided to use a two-step approach, by first dating crown-group *C. sordidula* using secondary age estimates on the outgroup divergences, and then using the estimated crown age to calibrate an intraspecific dating analysis. This approach allowed us to avoid dating a mixed phylogeny with inter- and intraspecific divergences.

To date the crown group node of *C. sordidula*, we trimmed the 90%-0.98-spruceup alignment to only include the three outgroup taxa *C. scutellaris*, *C. nigropilosa* and *C. longipilosa*, and two *C. sordidula* samples (BBX973 and BBX890) from two robustly supported clades encompassing the root node of the *C. sordidula* phylogeny. We used approximate likelihood to estimate divergence times as implemented in MCMCTree in PAML v.4.10.6 (Yang 2007) and secondary calibrations on the basis of the global *Crematogaster* phylogeny from Blaimer (2012). We hereby used median ages from three analyses in that study (Blaimer 2012), calibrating the *Crematogaster* crown age (root including the outgroups in our study) from 40.5 to 44.9 Ma, the most recent common ancestor (MRCA) of the *Orthocrema* clade with 25.6–30.2 Ma (including *C. sordidula*, *C. longipilosa*, and *C. nigropilosa*), and the divergence between *C. longipilosa* and *C. sordidula* to 16.9–19.2 Ma. Data for node calibrations are listed in Table S5. The analysis followed the protocol by Dos Reis and Yang (2019), with specifications of parameters similar to Ward and Blaimer (2022). We set up four independent runs for approximate likelihood calculations, using the independent rates model, the HKY85 model, and values for alpha and ncatG set to 0.5 and 5, respectively. We achieved convergence using burnin = 500,000, nsamples = 1,000,000 and samplefreq = 10. For evaluation of convergence statistics, we used TRACER v1.7.2 (Rambaut et al. 2018), checking for ESS values (> 200) and posterior densities.

Initially, we estimated the divergence times for the ingroup of *C. sordidula* on the basis of the full 90%-0.98-spruceup alignment (with outgroups pruned) in MCMCTree, following the same steps as described for dating the root of *C. sordidula* while using the results from this first analysis (mean: 3 Ma, 95% highest posterior density (HPD) interval > 0.6 < 5.5 Ma; Table S5) to calibrate the root node. However, this analysis produced ages for the deeper nodes that seemed implausibly old and estimated a younger root age for the ingroup compared to the estimates obtained when using more calibration points during the initial dating step.

We therefore ran a second dating analysis for ingroup *C. sordidula* using SNAPP version 1.6.1 (Bryant et al. 2012) implemented in BEAST version 2.7 (Bouckaert et al. 2014), following the method described by Matschiner (2022), again using the result of the first MCMCTree analysis to calibrate the root node (mean: 3 Ma, 95% highest posterior density (HPD) interval 0.6–5.5 Ma, Figure S6 and Table S5). For this analysis, we used the SNP alignment on the basis of the 14 previously delineated units with two samples each (one sample for Morocco (11), Table S4). A strict clock model was implemented, assuming that only closely related species are used in the analysis. We estimated mutation rates from the data ($u = 2.938$ and $v = 0.603$), leaving other priors at default settings. We removed the treeNodeSwapper operator to fix the tree topology to our previously estimated result and ran the analysis with 10 million iterations, sampling every 1000 steps. We performed the analysis twice with the same settings and combined the results of both runs and created a maximum clade credibility (MCC) tree using TreeAnnotator version 2.7.4 (Drummond and Rambaut 2007). Convergence was examined using TRACER version 1.7.2 (Rambaut et al. 2018). The SNAPP analysis yielded smaller HPD intervals and more credible results than the MCMCTree analysis, and we therefore consider those results better-supported and choose them for use in subsequent biogeographic analyses.

2.7 | Biogeographic Analyses

To infer the biogeographical history of *C. sordidula*, we estimated ancestral ranges using several models implemented in BioGeoBEARS (Matzke 2013) in R v.4.3.0 (R Core Team 2025), on the basis of the maximum credibility tree estimated in the SNAPP analysis. Lineage occurrence was categorised according to 9 areas: Morocco (M), Spain + south France (S), Italy (I), Adriatic Coast (D), Balkan Peninsula (P), Aegean region (A), Balkan Mainland (B), Türkiye (T), and Crete (C). Analyses used six biogeographical models specified in the package: DEC—Dispersal-Extinction Cladogenesis model; DIVALIKE—ML version of Dispersal Vicariance Analysis; BAYAREALIKE Bayesian biogeographical inference model, with and without an additional parameter (+J) for each of these models simulating the process of founder-event speciation (Matzke 2014). Given that samples assigned to the same lineage in the condensed dataset occur in up to three areas (Table S4), we set max_range_size = 3. All other parameters were left as default values. We used log-likelihood (LnL), Akaike information criterion (AIC), and sample-size corrected AIC (AICc) scores to compare each model's suitability to our data (Table S6).

2.8 | Extraction of Mitochondrial Data and Haplotype Reconstruction

To test whether mitochondrial haplotypes alone recover patterns congruent with our genome-wide UCE data or reveal signs of incongruence due to processes like introgression or incomplete lineage sorting, we also extracted up to 658 bp of the cytochrome oxidase I (COI) barcode region from our assembled contigs, using the phyluce_assembly_match_contigs_to_barcodes.py script from the PHYLUCE package. For this, we first downloaded a COI barcode sequence of a *C. sordidula* sample from the NCBI database as a reference (https://www.ncbi.nlm.nih.gov/nuccore/MT606315.1). After matching the contigs to the barcode reference, the resulting files containing multiple sequence fragments were divided using Splitfasta v1.0.0, and we kept the longest fragment. Then, sequences were concatenated and uploaded to BLAST (https://blast.ncbi.nlm.nih.gov) to verify the species identity and check for contamination. When BLAST results indicated a taxon other than *C. sordidula* for one of the sequences, those were removed from the COI dataset. All sequences ≥ 400 bp were aligned using MAFFT v.752 (Katoh et al. 2002), leading to inclusion of 63 taxa in the final "COI_Haplo" alignment. The alignment was further trimmed to 335 bp to exclude any missing data, which is required for haplotype reconstruction. Shorter barcodes (> 200 bp) perform comparably to full-length barcodes for species delimitation and species identification (Yeo et al. 2020). Therefore, no detrimental effects are expected for our analyses. However, the performance of short barcodes in intraspecific analyses may be less effective than the conclusions of Yeo et al. (2020) suggest.

The same nine biogeographic regions were defined for haplotype reconstruction as for biogeographic analysis. A trait matrix was constructed by assigning individuals to one of these regions on the basis of their collection locality and used as input for haplotype reconstruction in PopART (Leigh and Bryant 2015). Haplotype reconstruction was done using the integer neighbour-joining method (intNJ, $\alpha = 1$), as this method has been suggested as the best option for inferring haplotype networks for low-divergence datasets (Leigh and Bryant 2015). Several statistics were calculated in PopART, including Tajima's D statistic, nucleotide diversity, segregating sites, and parsimony-informative sites (Table S7).

2.9 | SNP Alignment Processing for PCA, ADMIXTURE and Genetic Differentiation

To obtain a dataset suitable for population genomic analyses, we assembled a second, new SNP alignment from the UCE data. We implemented a custom SNP-calling workflow that converts per-locus UCE alignments for both alleles per sample into a single multi-sample VCF file containing one high-quality SNP per locus. We used the output of the allele phasing from the PHYLUCE pipeline (see above) and a custom Python script (uce_to_vcf.py), which parsed each alignment, grouped allele pairs by specimen, and scanned nucleotide positions for variable sites. To avoid biases associated with linkage and to ensure comparability across loci, we extracted a single SNP per UCE locus, choosing a random polymorphic position in the alignment that contained only unambiguous bases (A, C, G, T). Because our SNPs are derived from targeted UCE loci rather than from a dense genome-wide SNP panel, we followed common UCE population-genomic practice of thinning to one SNP per locus and treating these markers as unlinked, rather than performing additional LD-pruning in sliding windows. Similar approaches have been used in several UCE-based population-structure and demographic studies (e.g., Winker et al. 2018; Erickson et al. 2021; Stiller et al. 2023; Naughton et al. 2024). Sites containing gaps or ambiguous characters in any allele of any sample were excluded from evaluation. For every retained

SNP, genotype calls were derived directly from the two phased haplotypes of each specimen. The reference allele was defined as the alphabetically first nucleotide present at that site across all samples, and all additional alleles were encoded as ALT variants in a multiallelic VCF format. Samples missing one or both alleles at a site were encoded as missing genotypes ("./."). The resulting file was exported as a valid VCFv4.2 dataset containing one SNP per UCE locus (2481 SNPs). The combined VCF containing all retained SNPs (one per UCE) was processed with PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/, Purcell et al. 2007). Each locus identifier was treated as an independent contig and accepted via the allow-extra-chr option. The VCF was converted to binary PLINK format. Basic quality statistics (allele frequencies, per-SNP and per-individual missingness, and Hardy–Weinberg equilibrium tests) were computed in PLINK. Quality filtering followed standard thresholds for population genomic datasets: minimum minor allele frequency MAF ≥ 0.05 and maximum per-SNP missingness GENO ≤ 0.05.

2.10 | PCA and ADMIXTURE Analyses

The binary genotype files were used as input for a PCA conducted in PLINK using default parameters. We calculated matrices for the first 20 PCs, resulting in the pca_results.eigenval and pca_results.eigenvec files. Population structure was inferred with ADMIXTURE v1.3 (Alexander et al. 2009) using the quality-filtered PLINK dataset. Analyses were run for $K = 1$–10 with 10-fold cross-validation. For each value of K, the Q-matrix (individual ancestry coefficients), P-matrix (allele frequencies), and cross-validation error were recorded. CV values reached a plateau at $K = 4$–6, which can be used as a first indicator for choosing the appropriate K. However, the results from PCA analysis and phylogenetic analyses, as well as biological context, were used additionally to identify the best-fitting K for this study.

2.11 | Pairwise Genetic Differentiation Between Regions

We analysed genetic differentiation and estimated gene flow among regions on the basis of Fst values using the quality-filtered dataset outputted from PLINK. Individuals were assigned to 10 geographical regions (M, S, I, D, P, B, A, C, T, Y) on the basis of sampling localities. Regions were defined similarly to the regionalization used for the haplotype reconstruction and biogeographic analysis (see above), except that we additionally split Sicily (Y) from the remaining Italian Peninsula (I) to test for genetic diversity of island lineages in comparison to mainland populations (Table S4). Sample sizes per region (n) ranged from 3 individuals (M) to 22 individuals (A and B). Genetic differentiation between populations was calculated using Weir and Cockerham's Fst in PLINK with default parameters, downstream analyses and data handling were performed in R, and F-statistics and population-level summaries were additionally calculated using the hierfstat package (Goudet 2005). To visualise genetic connectivity among regions, we constructed pairwise FST-based networks using region centroids derived from sampling coordinates. Connectivity edges were weighted as an inverse function of genetic differentiation ($1/[1 + FST]$), and network structure was summarised using a minimum spanning tree (MST) to represent the most parsimonious backbone of genetic connectivity across regions using the minimum total genetic differentiation, providing a parsimonious summary of relative genetic similarity by retaining only the lowest-distance connections and removing redundant edges. To test specifically for gene flow between two genetically distant lineages co-occurring on the Italian Peninsula, we assembled a second dataset and re-computed pairwise Fst values, this time with samples BBX1001 and BBX948 being in one population and the remaining Italian Peninsula samples in the other.

2.12 | Spatial Genetic Diversity

Genetic diversity was quantified using different standard population genomic parameters (Ho/He, F, and π values). To estimate genetic diversity, the full SNP matrix prior to filtering for one SNP per locus (200,862 SNPs) was used as input and analysed using the vcfR package in R (Knaus and Grünwald 2017). We compared all values between the 10 regions and tested for statistical significance using a Kruskal–Wallis test, Pearson's product–moment correlation, and Dunn Kruskal–Wallis multiple comparison in R using all_stats (vcfR package).

All analyses were performed either on an iMac Retina 5K with a 3,4 GHz Quad-Core Intel Core i5, a Linux high-performance cluster housed at the Museum für Naturkunde, Berlin, or a 64-core, 1TB RAM Linux server housed at the California Academy of Sciences, San Francisco.

3 | Results

3.1 | UCE Sequencing Results

The mean DNA sample concentration for the 126 samples was 0.91 ng/μL (< 0.05–7.37 ng/μL) after extraction and 16.92 ng/μL (< 5.0–42.7 ng/μL) for post-PCR libraries. After sequencing, we received an average number of 2,150,928 reads per taxon (range: 402,393–3,339,946 reads) per sample. From the assembled contigs, we captured, on average, 2247 UCE loci per specimen with a mean length of 954 bp. Additional details on sequence data statistics are provided in Table S2.

UCE loci were filtered using two completeness settings (90%, 80%) and subjected to trimming with Spruceup using three different cutoff values. After comparison of the phylogenetic results, we found that trimming with a cut-off value of 0.98 proved optimal for removing long terminal branches, whereas trimming with a cutoff of 0.97 and 0.95 resulted in loss of informative sites (i.e., internal branches getting shorter) and indicated overtrimming. Thus, we chose to proceed with the 90% completeness matrix with 0.98 spruce-up trimming as the most complete dataset for dating and biogeographic analyses. The final alignment has a length of 1,942,568 bp with 18% missing data, containing 234,088 (12.1%) variable sites and 53,863 (2.8%) parsimony-informative (PI) sites. Alignment characteristics were calculated for all matrices with AMAS and are shown in Table S3. Spruce-up trimming mostly affected long branches with high proportions of missing data, low locus capture, and short contigs (Tables S2, S3 Figures S1–S5).

3.2 | Phylogenetic Structure *of C. sordidula*

We inferred phylogenetic relationships of *C. sordidula* populations on the basis of a concatenated ML and a coalescent approach and compared tree topologies across analyses. We observed similar patterns for all concatenated trees for the main clades, comparing support values using the following categories. Full support: ultrafast bootstrap (ufBS) and SH-aLRT = 100%, and local posterior probability (LPP) = 1; high support: ufBS = 95%–99%, SH-aLRT = 80%–99%, LPP = 95–99; low support: ufBS < 95%, SH-aLRT < 80%, LPP < 95. Some minor changes in topology can be observed in areas with low support when comparing the results using different completeness matrices and spruceup cut-off values, and employing partitioning vs. no partitioning (Figures S1–S5). These have no relevance to the general patterns observed for the main clades; therefore, we focus on the phylogenetic hypotheses on the basis of the concatenated 90%-0.98-spruceup matrix in the following (Figures 2, S1).

Both concatenated ML and coalescent analyses indicated four main clades with full support, which we named by region: (i) Western Mediterranean clade (WM), including samples from southern France, Spain, Morocco, and the Tyrrhenian coast of Italy; (ii) Eastern Aegean region clade (EA), including samples from coastal and southern Türkiye and eastern Aegean islands; (iii) Anatolia-Türkiye clade (AT), encompassing northern, eastern and central Türkiye and part of Bulgaria and Greece (Eastern Macedonia and Thrace); (iv) Central-Eastern Mediterranean clade (CEM), including samples from the Adriatic coast, the Balkan Peninsula, Euboea and Central Aegean islands (Cyclades) and Crete. In all analyses, the WM clade is estimated as the sister lineage to all remaining taxa with full support. Among the three remaining clades, representing together a larger Eastern Mediterranean clade, the EA clade was identified as sister to both the AT clade and the CEM clade. These results are shown in Figures 2 and S1 and were fully supported by all analyses.

Among the WM clade, a sample from Morocco (BBX973) is fully supported as the sister lineage to all others in the concatenated analysis (Figure 2). In the coalescent analysis, a sample (BBX1094) from southern France (Pyrenees) is supported in this position (Figure S1), but this sample had significantly less locus coverage (1258 loci vs. an average of 2247 loci across the dataset, Table S2). Our phylogenetic reconstructions revealed two subclades inside the WM clade in the concatenated analysis, but the coalescent topology only partially agrees with this result, and there are several rearrangements of lineages (Figure S1). The EA clade is estimated to be the sister group to all other eastern Mediterranean lineages in both coalescent and concatenated analyses with full support. Within the clade, both analyses agree on the relationships of major lineages, but there are some rearrangements of shallow nodes. The AT clade consists of lineages from the far east of the Mediterranean area at the border with the Asian continent. This clade is sister to the CEM clade and has a ladderized structure, lacking larger subgroupings of lineages.

A major part of samples included in our analyses is nested inside the CEM clade. One noteworthy finding fully supported by all concatenated results is that all five samples from Crete and the neighbouring island Karpathos form a monophyletic group (Crete clade, highlighted in green in Figure 2). The Crete clade is estimated as the sister group to most other lineages in the CEM clade, exclusive of two samples from Türkiye (BBX1129) and Bulgaria (BBX942). However, although the Crete clade is fully supported in the coalescent tree (Figure S1, LPP = 1) as well, its position is not resolved (LPP = 0.17). The remaining relationships of taxa within the CEM clade are generally not well supported, with the exception of some shallow nodes.

3.3 | Divergence Dating and Biogeographical History

Our dating analysis performed in MCMCTREE estimated the most recent common ancestor (MCRA) for Mediterranean *C. sordidula* to be of Late Pliocene origin (mean: 3 Ma, 95% HPD: 0.6–5.5 Ma, Table S8 and Figure S6). These values were then used as calibration to infer the divergence ages for the intraspecific phylogeny of *C. sordidula* in SNAPP on the basis of a reduced clade-level dataset. This analysis estimated the calibrated MRCA for all *C. sordidula* lineages as Late Pliocene, around 3.2 Ma (Table S8 and Figure 3 node 15, 95% HPD: 2.89–3.49 Ma). For the WM clade, the MRCA is dated at 2.4 Ma (95% HPD: 2.2–2.66 Ma, node 26), and its two subclades diverged only slightly later (2.3 Ma, node 27).

The MCRA of the Eastern Mediterranean group (clades EA, AT, CEM; Figure 3) is estimated with an age of 3.15 Ma (95% HPD: 2.86–3.45, node 16) in the Late Pliocene, when the EA clade diverged from the remaining Eastern Mediterranean group. The remaining two major clades (AT, CEM) diverged around 3 Ma (node 17). The MRCA of the AT clade is estimated around 2.7 Ma (node 18), whereas the MRCA of the CEM clade is estimated slightly younger at around 2 Ma (node 20). Deeper-level diversification events within the major clades took place between 1.2–1.7 Ma, a period that overlaps with the onset of the Mid-Pleistocene Transition (MPT, Herbert et al. 2023), when glacial–interglacial contrasts and sea-level amplitudes progressively intensified.

The comparison of all six models of ancestral area estimation with BioGeoBEARS recovered DIVALIKE+J as most suitable for our data on the basis of the highest LnL and AICc scores (LnL = −43.9, AICc = 93.8; Table S6). Since the founder event speciation (+J) model in BioGeoBEARS has been criticised (Ree and Sanmartín 2018), we compare the results of the DIVALIKE+J analysis with the best-scoring model without the +J parameter (DIVALIKE; LnL = −48.4, AICc = 100.7; Table S6). Because of the relatively short distances between regions in the Mediterranean area and frequent reconnections of regions during low water levels, a founder event speciation model may be less appropriate for our data. In the following, the most important deviations between the DIVALIKE+J and DIVALIKE analyses are outlined, but we present the results of the DIVALIKE analysis in Figure 3. Detailed results of both analyses can be found in Tables S6 and S8.

Both models suggest that the MRCA of *C. sordidula* had a widespread distribution across the southern Mediterranean area, possibly extending from Türkiye across Northern Africa to Morocco, albeit with low probabilities of $p = 0.37$ (DIVALIKE; node 15,

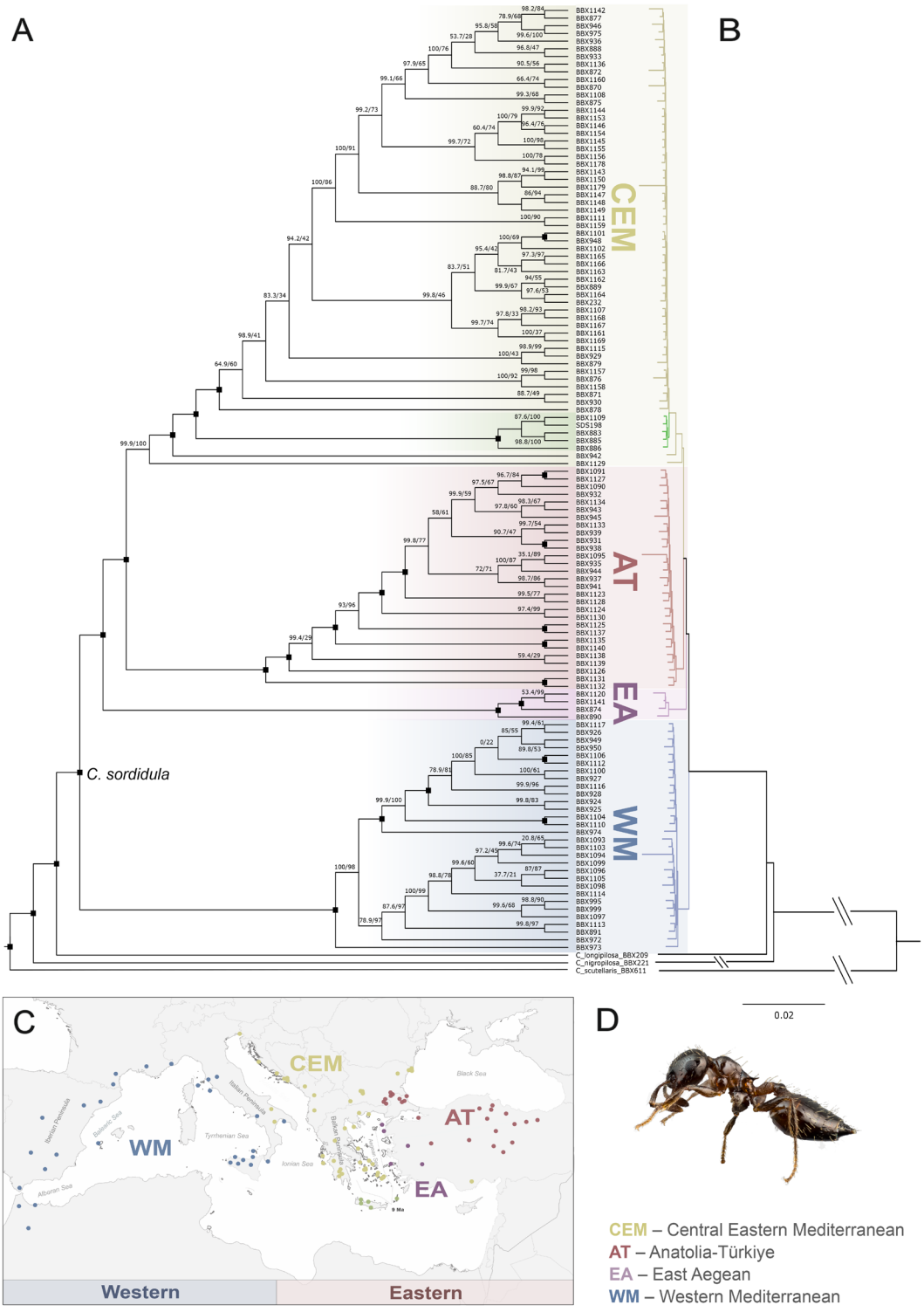

FIGURE 2 | Phylogeography of *Crematogaster sordidula* in the Mediterranean region. (A) Cladogram on the basis of analysis of the 90% completeness matrix, including partitioning and spruceup trimming with 0.98 cut-off. Highlighted on the tree are four clades as indicated in the legend: (CEM) Central-Eastern Mediterranean (yellow), including populations occurring on Crete and Karpathos (green); (EA) Eastern Aegean region (purple); (AT) Anatolia-Türkiye region (red); and (WM) Western Mediterranean (blue). The Crete clade is marked by green shading within the CEM clade. Branch support values are SH-aLRT/ultrafast bootstrap (ufBS); black squares indicate 100% support by both metrics. The phylogram in (B) illustrates the four major clades plus the Crete clade with branch colours corresponding to shading in (A). Branches of outgroup taxa were shortened as indicated by the intermittent lines; the scale bar represents nucleotide substitution per site. (C) Map of the Mediterranean region including all sample sites. Clade distribution areas are marked by colours and abbreviations according to the phylogeny, the east–west partition is indicated below. Thin dashed line denotes the mid-Aegean trench (9 Ma) in the Aegean Sea (D) Close-up of *C. sordidula* in lateral view. Photo credit: Jody H. Voges.

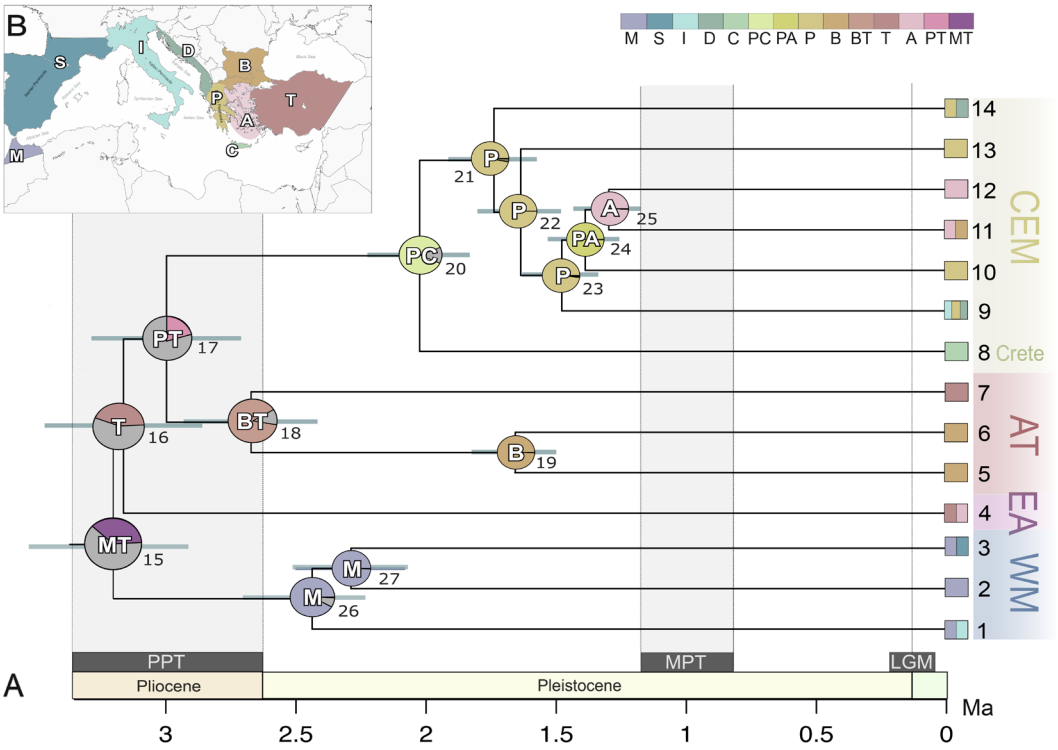

FIGURE 3 | (A) Ancestral area estimation of *Crematogaster sordidula* in the Mediterranean region. The biogeographical analysis was performed under the DIVALIKE model in BioGeoBEARS using the maximum clade credibility tree resulting from SNAPP analysis on the basis of 14 delineated phylogenetic units, here shown as numbered terminals. Probabilities for ancestral-area reconstructions are given as percentages in the text and are depicted as pies; nodes are numbered, and complete data for range probabilities are provided in Table S8. Pie states are: M = Morocco; S = Spain + southern France; I = Italy; D = Adriatic Coast; P = Balkan Peninsula; A = Aegean region; B = Balkan mainland; T = Türkiye; C = Crete and Karpathos; MT = Morocco–Türkiye; PT = Balkan Peninsula–Türkiye; PC = Balkan Peninsula–Crete; PA = Balkan Peninsula–Aegean region; BT = Balkan mainland–Türkiye; grey = all other states (combined). Clades are: WM, Western Mediterranean; EA, Eastern Aegean; AT, Anatolia–Türkiye; CEM, Central-Eastern Mediterranean. The timescale is shown in 0.5 Ma increments; major geological and climatic intervals are indicated below the timescale. The Pliocene–Pleistocene Transition (PPT), the Mid-Pleistocene Transition (MPT), and the Last Glacial Maximum (LGM) are marked. (B) Map of the Mediterranean region with areas coloured according to the biogeographical analysis.

Table S8, Figure 3) and $p = 0.21$ (DIVALIKE+J; Table S8, Figure S7). On the basis of the reconstructions, the main split of Eastern and Western Mediterranean lineages within *C. sordidula* happened during the beginning of the Pliocene–Pleistocene Transition (PPT, 3.2 Ma). For all eastern Mediterranean populations of *C. sordidula*, including the EA clade, Türkiye is the most probable ancestral area (Table S8, Figures 3 and S7; T, $p = 0.30$–0.44, node 16). The DIVALIKE reconstruction indicates that following the divergence of the EA clade, the remaining Eastern Mediterranean populations expanded their distribution to the Balkan Peninsula (Table S8 and Figure 3; PT, $p = 0.20$, node 17). In contrast, DIVALIKE+J supports the distribution of the MRCA of the AT + CEM clades still as restricted to Türkiye (Table S8 and Figure S7; T, $p = 0.30$, node 17). However, both scenarios received very low probabilities.

In the DIVALIKE scenario, the ancestral area for the AT clade is estimated as Balkan mainland and Türkiye (Table S8 and Figure 3; BT, $p = 0.88$, node 18). Vicariance in the early Pleistocene within this clade then led to the further diversification of lineages from the broader ancestral area to lineages restricted to Türkiye and the Balkan Peninsula (B, $p = 1$, node 19). However, the DIVALIKE+J model suggested that the MRCA of the AT clade still resided in Türkiye (Table S8 and Figure S7; T, $p = 0.46$, node 18), indicating a dispersal event from Türkiye to the Balkan mainland (B, $p = 1$, node 19) in the early Pleistocene.

For the CEM clade, both analyses suggest an MRCA from a broader region including the Balkan Peninsula and Crete (Table S8, Figures 3 and S7; PC, node 20; DIVALIKE: $p = 0.83$; DIVALIKE+J: $p = 0.33$). However, in the DIVALIKE+J model, more restricted ancestral areas of either the Balkan Peninsula or Crete are comparably likely (P, $p = 0.31$; C, $p = 0.32$; node 20). Either way, a lineage now restricted to Crete and Karpathos appears to have diverged relatively early, around 2 Ma. Both models suggest diversification of the CEM clade mainly on the Balkan Peninsula (Table S8, Figures 3 and S7; P, $p = 0.93$–1, nodes 21–23). The DIVALIKE model estimates a dispersal from the Balkan Peninsula to the islands of the Aegean region around 1.4 Ma (Table S8 and Figure 3; PA, $p = 0.99$, node 24), whereas the DIVALIKE+J model suggests this dispersal event to have occurred slightly later, around 1.3 Ma (Table S8 and Figure S7; A, $p = 0.97$; node 25). From the islands of the Aegean Sea, one lineage appears to have dispersed back to the Balkan mainland during the later period of the Pleistocene.

The WM clade established in the western Mediterranean region in the early Pleistocene, and both models estimate that lineages have dispersed from northern Africa (Table S8, Figures 3 and S7; M, $p = 0.87$–0.92, node 26) to Italy sometime after 2.4 Ma, and subsequently from Morocco to Spain (M, $p = 0.98$–0.99, node 27).

3.4 | Haplotype Network Reconstruction and Analyses

The final COI_Haplo matrix and network contain eight out of nine regions, because no COI data was obtained for samples from Morocco (Figure 4, Table S9). We found 23 different COI haplotypes, which can be grouped into five haplotype groups (WM, EA, AT, CEM, and Crete; Figure 4, Table S9). The COI data used for this analysis contain a nucleotide diversity of pi = 0.074072, 74 segregating sites and 66 parsimony-informative sites (Table S7). Group WM is the most distant, separated by 16 single nucleotide polymorphisms (SNPs) from the closest estimated ancestral haplotype, and 16 + 8 + 9 SNPs (33 SNPs total) separating it from the most closely related haplotype group (CEM).

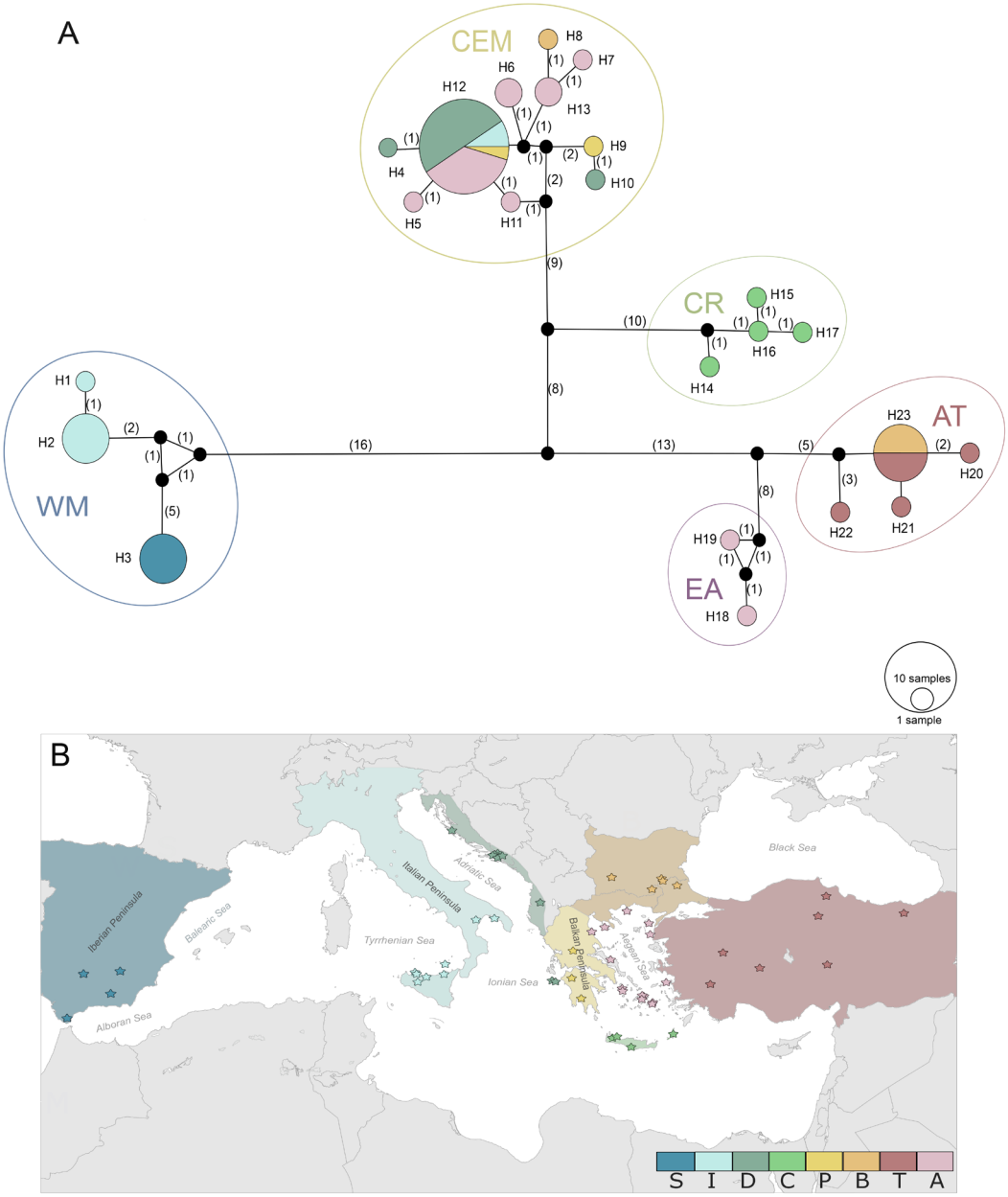

FIGURE 4 | Haplotype network analysis on the basis of COI. (A) Haplotype network on the basis of the integer neighbour-joining method and a 335 bp long fragment of *cytochrome oxidase 1* (COI) for the reduced dataset including 63 samples from the original dataset. Sequenced haplotypes are represented by circles and named from H1 to H23 according to Table S9; their size is proportional to the frequency with which the haplotype occurs in the dataset. Numbers on the branches indicate the number of single-nucleotide polymorphisms between haplotypes; black circles represent inferred ancestral haplotypes not represented by the dataset. Haplotype groups are marked by coloured circles and abbreviations named after phylogenetic clusters (WM = Western Mediterranean EA = Eastern Mediterranean, AT = Anatolia-Türkiye, CEM = Central Eastern Mediterranean and CR = Crete and Karpathos). Colours show the region of occurrence for each haplotype. (B) Map of the Mediterranean region with sample sites used in the haplotype network indicated by stars. Regions are coloured according to the bioregions indicated in the legend; S = Spain + southern France, I = Italy, D = Adriatic Coast, P = Balkan Peninsula, A = Aegean region, B = Balkan Mainland, T = Türkiye, C = Crete and Karpathos. Moroccan samples were excluded from the dataset because of the absence of retrievable COI sequences.

Group WM contains three haplotypes from two regions (Spain + southern France and Tyrrhenian coast). H1 is unique for one sample from the Tyrrhenian coast of Italy (BBX1110), whereas H2 is shared by the rest of all Tyrrhenian coast samples. Samples from Spain and southern France all share the same haplotype (H3).

The haplotype group EA contains two haplotypes (H18, H19) from the east Aegean region (samples BBX1141 and BBX874). The group is separated by 8 + 5 SNPs (13 SNPs total) from group AT, and 8 + 13 SNPs (21 SNPs total) from the ancestral haplotype shared with groups WM, CEM and Crete. Group AT is separated from the shared ancestral haplotype with group WM, CEM and Crete by 13 + 5 SNPs (18 SNPs total) distance and has four haplotypes, three of which are unique (H20, H21, H22) and occur in Türkiye. The fourth haplotype in this group, H23, is shared by eight samples, half of them from Türkiye, and half from the Balkan mainland.

The haplotype group CEM contains samples from the Balkan Peninsula and Aegean region, Adriatic Coast, and one sample from the Balkan mainland. With 10 different haplotypes, it has the highest haplotype diversity in our study. Seven of the 10 haplotypes are unique for one individual each (H4, H5, H7–11). Three haplotypes are shared among samples (H6, H12, H13). H6 is unique for two individuals from the Aegean region (BBX1144, BBX1153), whereas H12 is shared among 22 samples from three regions (Aegean region, Balkan Peninsula, and Adriatic coast). One haplotype (H13) occurs in two samples from the Aegean region (BBX1136, BBX877).

The fifth group contains four unique haplotypes from Crete and Karpathos, differing from each other by one or two SNPs. This haplotype group is segregated by 10 + 9 + 1 SNPs (20 SNPs total) from the most closely related haplotype from the Aegean region (H11).

3.5 | PCA and ADMIXTURE Analysis

The first three principal components explain 75.61% of all variation for the dataset. PC1 explains 49.77% variance, the largest source of variation in the dataset. PC2 explains 22.12%, and PC3 3.72% (Figure 5D). PC4–PC20 explain progressively smaller portions (0.7%–3%) of the total variance; the proportion of contribution of different clusters to the PCs is visualised in Table S10. Along PC1, populations cluster into western and eastern lineages; adding PC2 separates CEM and AT clades and reveals some samples shared between the three main clusters, including a group of four samples representing the EA clade.

Our results show a plateau of CV values for $K = 3$–6, with $K = 3$ receiving the best score followed by $K = 5$ with a non-significant difference (Table S11). We decided on $K = 5$ as the best fit to our data, because of better alignment with results from the phylogenetic and haplotype reconstructions and the PCA, while explaining more details of the ancestral history of *Crematogaster sordidula* without oversplitting populations. In the following, we describe the results for $K = 5$; however, the results for $K = 3$, 4, 6, and 7 are also presented in the Figures S8–S11. At $K = 5$, ADMIXTURE produced five pure clusters (C1–C5) and several admixed clusters combining two ancestry components. ADMIXTURE assignments showed a high degree of correspondence with the main clades defined on the basis of the phylogenetic reconstructions (Figure 5A). All WM samples were assigned purely to cluster C3, all EA individuals purely to C1, and the majority of AT samples (24) purely to C4. Individuals assigned to the CEM clade—including all Crete clade samples—occurred as pure samples in clusters C2 and C5, and as admixed samples in their corresponding mixed clusters C2_C5 and C5_C2. Only a single pure sample did not match its phylogenetic assignment, a CEM individual (BBX1095) placed in cluster C4 instead of the expected C2/C5 cluster group. Most admixed samples (42) belonged to the reciprocal mixed clusters C2_C5 and C5_C2 and are assigned to the CEM clade on the basis of the phylogeny. Additionally, five AT clade samples were assigned to admixed clusters C4_C1 and C4_C2, and one CEM sample (BBX942) to C4_C5. Crete clade individuals were recovered as pure samples in cluster C2 but did not form a distinct ADMIXTURE cluster under $K = 5$, only at $K = 7$ (Figure S11). No WM or EA samples were assigned to admixed clusters. All admixture clades for $K = 2$–10 are shown in Table S12.

3.6 | Pairwise Genetic Differentiation and Patterns of Genetic Connectivity

Pairwise Fst values among the 10 regions revealed a clear population structure (Figure 6) mostly consistent with the main phylogenetic clades. Hierarchical clustering highlighted three major genetic groupings: (1) a cluster containing I, M, S and Y, with M, S and Y forming a more closely related group; (2) an intermediate cluster of B and T; and (3) a cluster including A, C, D and P (Figure 6). Overall differentiation ranged from zero to 0.44, indicating generally low to moderate genetic divergence. Populations in regions I, M, S and Y exhibited very low Fst values among them (0.00–0.05), suggesting minimal differentiation and substantial shared genetic variation. Similarly, the second (B and T) and third (A, P, D and C) clusters also showed no or minimal internal differentiation. In contrast, regions M, S and Y are strongly differentiated from A, P, D and C, with Fst values typically ranging from 0.38 to 0.44, representing the highest observed levels of divergence. The greatest differentiation exists between populations from clusters M and C (Fst = 0.44). In contrast, regions B, A and I show only moderate differentiation to regions outside their clusters (Fst = 0.15–0.28), indicating more recent genetic exchange or shared ancestral history for populations occurring in this region (Figures 6 and S12).

Overall, the pairwise Fst tests between populations revealed a distinctly structured west–central–east pattern of genetic differentiation across the Mediterranean basin. The strongest genetic connectivity (Fst < 0.02) is concentrated within the main clades WM (regions I, M, S, Y), CEM (including EA; regions A, C, D, P), and AT (regions B, T), with little gene flow between them. The greatest level of differentiation hereby occurs between WM (M, S, Y) and CEM (A, C, D, P) populations, whereas the AT clade (B, T) shows intermediate levels of separation, being closest to the CEM clade while still maintaining a distinct identity. Populations occurring on the two large islands, Crete (C) and Sicily (Y), both show high gene flow within their major clades (CEM and WM, respectively) but limited gene flow outside of

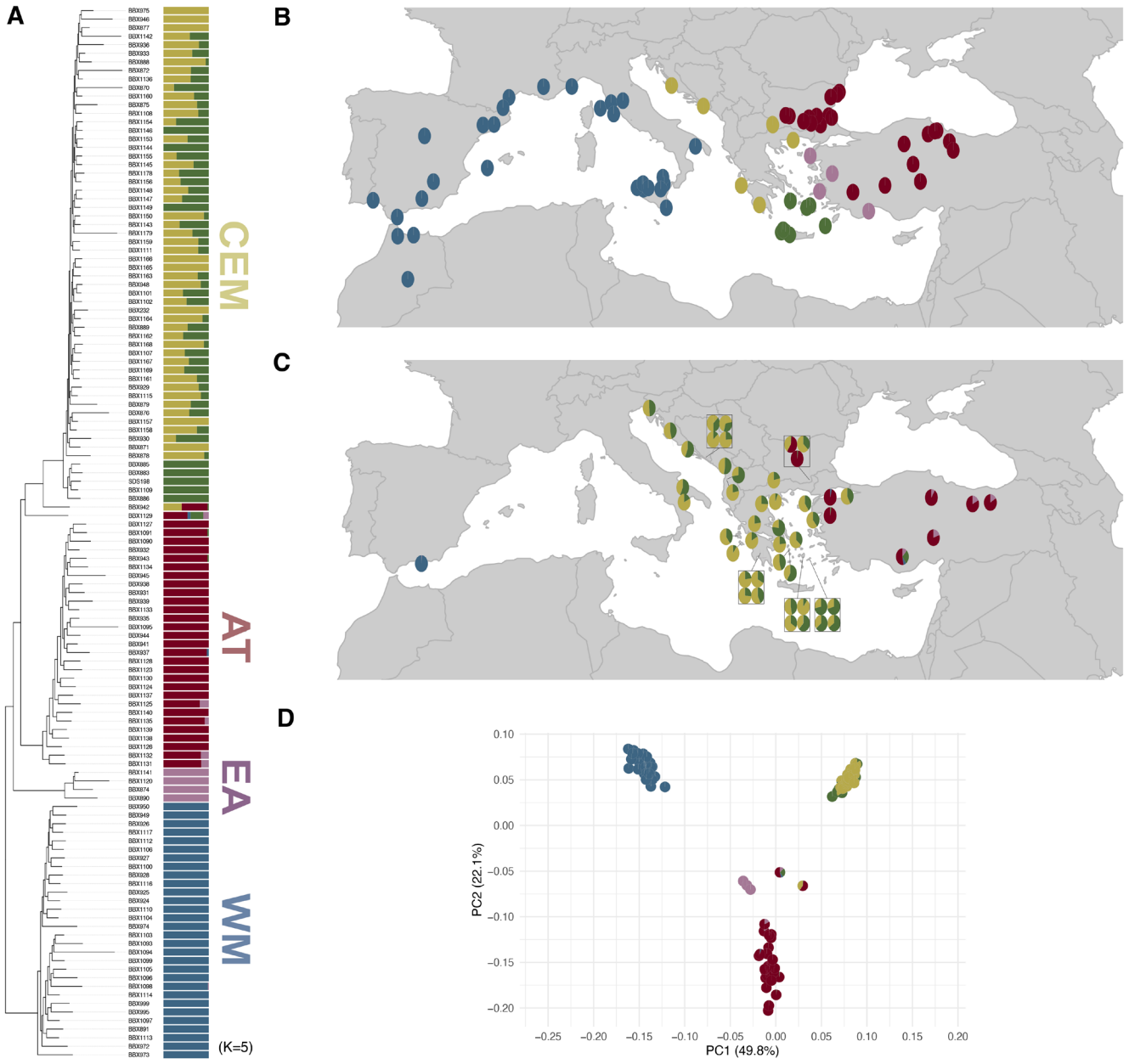

FIGURE 5 | Population genetic structure of *Crematogaster sordidula* inferred from ADMIXTURE and PCA analyses and visualised in a geographic context. (A) ADMIXTURE results for the best-fitting model ($K = 5$; Table S12), with individual ancestry proportions shown as horizontal bars. Individuals are ordered according to the phylogenetic relationships (Figure 2). Major clades identified in the phylogenetic analyses are indicated alongside the plot: WM (Western Mediterranean), EA (Eastern Aegean), AT (Anatolia–Türkiye), and CEM (Central-Eastern Mediterranean). (B) Geographic distribution of ADMIXTURE ancestry proportions ($K = 5$) across the Mediterranean region. Pie charts at sampling localities represent individual ancestry components, coloured according to the five inferred clusters: C1 = pink, C2 = green, C3 = blue, C4 = red, C5 = yellow. (C) Geographic distribution of admixed individuals under $K = 5$ across the Mediterranean region; (D) Principal component analysis (PCA) of all individuals on the basis of 2481 SNPs. The first two principal components are shown (PC1 = 49.77% and PC2 = 22.12%) representing 71.89% of the total variance.

them. Only populations on islands in the Aegean region (A) exhibit moderate levels of gene flow with populations outside their own cluster (A–B Fst = 0.15), underscoring the role of the Aegean region as a genetic melting pot within the eastern Mediterranean (Figures 6 and S12).

Since our phylogeny and the ADMIXTURE analysis showed that two samples on the Italian Peninsula are genetically closer to samples from the Balkan Peninsula (CEM clade), we also performed pairwise Fst tests adding a subdivision between the two distinct lineages on the Italian Peninsula (Figures S13 and S14). The Fst values highlight that the two aberrant samples cluster together with other populations from the eastern Mediterranean (CEM clade), whereas the remaining Italian populations are more closely connected to the western Mediterranean lineage. No gene flow was inferred between these two distinct populations (Fst = 0.39) on the Italian Peninsula, nor between the western Italian Peninsula population and any eastern Mediterranean

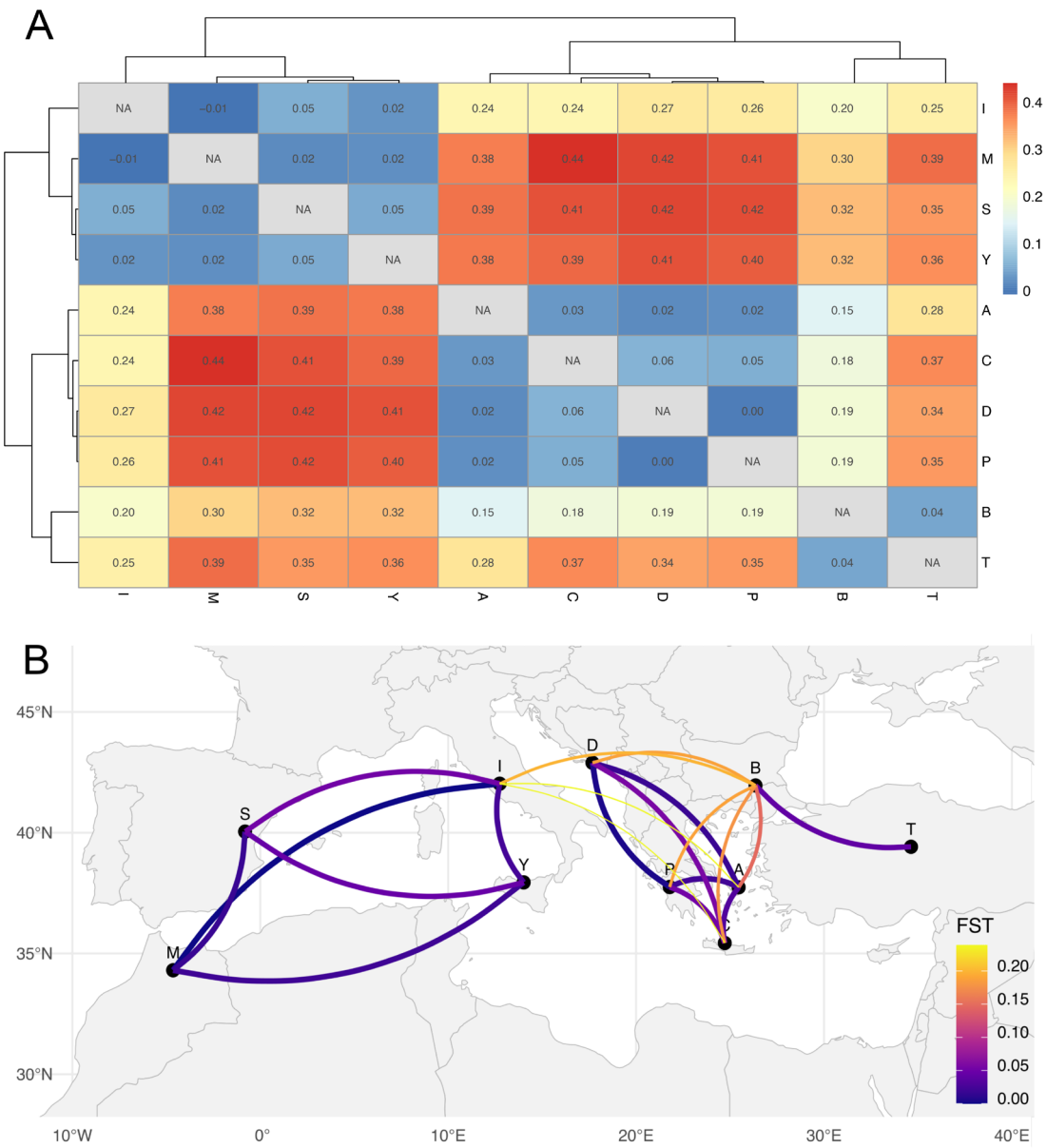

FIGURE 6 | Pairwise genetic differentiation and spatial patterns of genetic connectivity among 10 regions for Mediterranean populations of *Crematogaster sordidula*: (A) heatmap of pairwise Fst values among regions, showing the full range of genetic differentiation. The colour scale spans the complete empirical range of observed Fst values (0–0.44), with blue indicating low Fst values (low genetic differentiation), yellow intermediate values, and red high Fst (high genetic differentiation); (B) geographic representation of genetic connectivity inferred from pairwise Fst. Connections between regions are shown for low to moderate Fst values (0–0.2) to emphasise regions of higher inferred connectivity. Line colour transitions from dark blue, indicating the lowest Fst values and highest connectivity, to purple, red, and finally yellow, indicating progressively higher Fst values and lower inferred connectivity. Regions are: A = Aegean region, B = Balkan Mainland, C = Crete and Karpathos, D = Adriatic Coast, I = Italy, M = Morocco, P = Balkan Peninsula, S = Spain + southern France, T = Türkiye, Y = Sicily.

population. This finding supports the existence of two genetic lineages on the Italian Peninsula and emphasizes the reciprocal genetic isolation between eastern and western populations.

The minimum spanning tree (MST) constructed from the pairwise FST matrix highlights the core edges required to connect all populations with the smallest cumulative genetic distance (Figure S12B). The MST emphasizes the strong eastern–western split between populations and a subclustering of AT (B, T) and CEM + EA (A, P, D and C) populations, confirming the results from the remaining analyses. The western populations S, Y and M connect to the network through low-Fst edges to I, respectively, forming a western branch that mirrors their close genetic affinity observed in the Fst heatmap and hierarchical clustering (Figure 6). CEM + EA and AT populations (A, C, D, P and B, T, respectively) are more closely connected in the MST network, as observed in hierarchical clustering where they are two subclusters within a larger cluster.

3.7 | Spatial Distribution of Genetic Diversity

Overall, the genetic diversity estimates (Table 1) revealed marked differences in genetic variation among regions. Nucleotide

TABLE 1 | Spatial distribution of genetic diversity across Mediterranean regions. Mean genetic diversity statistics for *Crematogaster sordidula* populations across 10 Mediterranean regions. Genetic diversity metrics include H_o = observed heterozygosity; H_e = expected heterozygosity; F = inbreeding coefficient; and π = nucleotide diversity. Values represent regional means.

Region	H_o	H_e	F	π
A = Aegean region	0.22	0.25	0.15	0.25
B = Balkan Peninsula	0.33	0.33	0.01	0.33
C = Crete and Karpathos	0.18	0.15	−0.20	0.15
D = Adriatic Coast	0.18	0.17	−0.03	0.17
I = Italian Peninsula	0.22	0.22	0.01	0.22
M = Morocco	0.53	0.30	−0.77	0.30
P = Balkan Peninsula	0.23	0.24	0.04	0.24
S = Spain and southern France	0.30	0.34	0.11	0.34
T = Türkiye	0.18	0.18	0.00	0.18
Y = Sicily	0.21	0.21	0.00	0.21

diversity (π) differed significantly among regions (Kruskal–Wallis test (χ^2 *N* individuals = 123, regions = 10, SNPs = 239,722, $p < 2.2 \times 10^{-16}$); Table S13). Post hoc Dunn tests revealed significant pairwise differences between all regional comparisons, indicating strong and consistent regional structuring of genetic diversity. Expected heterozygosity (H_e) and nucleotide diversity (π) showed highly similar regional patterns, reflecting their shared dependence on allele frequencies. Regions B, M, and S exhibited the highest values of observed heterozygosity (H_o; 0.325–0.526) and π (0.298–0.340), consistent with genetically diverse populations and relatively large effective population sizes and higher historical connectivity. In contrast, regions C, D, and T showed the lowest H_o (0.178–0.183) and π (0.151–0.184) values, consistent with reduced genetic diversity potentially reflecting smaller effective population sizes, increased isolation, genetic drift, or past demographic bottlenecks. Inbreeding coefficients are low and generally close to zero across all regions, indicating no strong evidence of inbreeding.

4 | Discussion

The Mediterranean region is composed of numerous isolated islands, peninsulas, and mountains, and this topography has a profound influence on Mediterranean species diversity and distributions (Blondel et al. 2010). We examined the phylogeographic structure of *C. sordidula* in the Mediterranean region with a focus on the Balkan, Aegean, and adjacent regions. We estimated a population-level phylogeny, considering both concatenated and coalescent approaches. This framework was then used to perform a dating analysis and reconstruct the ancestral areas for the main intraspecific lineages. We additionally inferred the population structure of *C. sordidula* by calculating haplotype networks, admixture, and regional genetic diversity and differentiation. In the following, we discuss hypotheses for the dispersal and diversification of *C. sordidula* on the basis of our results under consideration of the complex climatic and geological history of the Mediterranean region.

4.1 | Origin and Early Diversification of *C. sordidula*

Our study supports the origin of *C. sordidula* somewhere in the southern Mediterranean region in the Late Pliocene. Given the placement of Moroccan samples in the phylogeny, it seems probable that *C. sordidula* originated in northern Africa. However, expanded sampling in this region would be necessary to confirm this hypothesis, including in eastern Algeria and Tunisia, whose biogeography is distinct from Morocco and whose ant faunas have connected with Italy (e.g., Schifani et al. 2022, 2020). We found a distinct population structure for *C. sordidula* with genetic clusters strongly reflecting their geographic distribution, suggesting equally complex dispersal patterns as for other Mediterranean-wide distributed taxa such as angiosperms, beetles, butterflies, centipedes, reptiles and spiders (Hammoud et al. 2021; Kornilios et al. 2019; Mas-Peinado et al. 2018; Trichas et al. 2020). As expected, we inferred a split between western and eastern Mediterranean populations that most likely occurred between 3.0 and 3.2 Ma, coinciding with the onset of the PPT. This east–west distribution pattern has been described in many taxa beyond ants, for example, in insects such as cave crickets (Allegrucci et al. 2011), grasshoppers (Noguerales and Ortego 2022) and flies (Martinez-Sañudo et al. 2024), and also in sister species of dominant Mediterranean tree genera (*Pinus*, *Juniperus* and *Quercus*) or birds (*Oenanthe*, *Sitta* and *Falco*) (Blondel et al. 2010). This is likely a result of allopatric lineage diversification and speciation because of isolation of the eastern and western Mediterranean biota over the last epochs (Blondel et al. 2010), when glaciation patterns in mountainous regions, especially in the Alps and Apennines during cooler paleoclimates, acted as a barrier to dispersal (Menchetti et al. 2021).

The Pliocene epoch ended with the onset of Northern Hemisphere glaciation and the commencement of a prolonged cooling trend (Khélifi et al. 2014). During Pleistocene climatic oscillations, Mediterranean taxa adapted to a warmer climate were impacted by the cold periods of the glaciations; for example, some butterfly species retreated southwards during glacial periods (Kühne et al. 2017). Particularly the glaciations in the Early Pleistocene highly influenced the distribution patterns of Mediterranean taxa (Hewitt 2001; Lukić et al. 2021; Taberlet et al. 1998; Taberlet and Bouvet 1994). The southern Mediterranean peninsulas and islands, characterised by high habitat diversity, are candidates as Pleistocene refugia for various taxa and likely played a crucial role in insect diversification, as evidenced by high rates of endemism (e.g., Crete, Sicily, Iberia, and the Balkan Peninsula; see Blondel et al. 2010).

Patterns of genetic diversity across regions can provide insights into locations of glacial refugia by identifying old genetic diversity that was preserved in a particular geographical area during that time (Feliner 2011). In our dataset, the Balkan mainland, Iberia, and Morocco (regions B, S, and M) exhibit the highest levels of genetic diversity, with elevated heterozygosity and nucleotide diversity as well as near-zero or moderately negative inbreeding coefficients. Such features are characteristic of long-term demographic stability, large effective population sizes, and

retention of ancestral variation, making these regions the strongest candidates for refugial areas during the Pleistocene. By contrast, Crete, the Adriatic coast, and Türkiye show markedly reduced heterozygosity and lower nucleotide diversity typically associated with more isolated populations shaped by reduced genetic connectivity and potentially affected by genetic drift or population bottlenecks.

4.2 | Diversification of the Western Mediterranean Clade

The western lineage was inferred with an origin in Morocco and restricted to the western Mediterranean during the onset and intensification of the Northern Hemisphere Glaciation (2.4–3.0 Ma). We distinguished two larger phylogenetic subclades within this lineage, one restricted to the Tyrrhenian coast of Italy and the other to southern France and Spain. However, these two subclades were not equally well supported in all phylogenetic analyses and our population genomic analyses indicated high levels of gene flow and connectivity between them, whereas admixture only inferred these as separate clusters at $K \geq 6$. The coalescent analysis inferred some of the samples from the southern France + Spain subclade as sister to the entire western clade instead of the samples from Morocco. Overall, a slightly lower number of UCE loci and shorter length of fragments in both involved groups of taxa are likely the cause for discrepancies between concatenated and coalescent analyses (e.g., Hosner et al. 2016), and results of population genomic analyses may also be impacted by the level of missing data in these samples.

The genetic divergence between the two subclades of the western lineage is estimated during an interglacial time, most likely with two dispersal events from northern Africa to the Iberian and Italian Peninsula, respectively. Dispersal from Africa to Iberia has been shown for other invertebrates, such as butterflies (Habel et al. 2008) and spiders (Korba et al. 2022). The straits of Sicily and Gibraltar were established during the Pliocene and generally promote allopatric diversification of arthropod lineages (Gantenbein and Largiadèr 2003). Yet, dispersal across the strait of Gibraltar has been inferred even for species with limited dispersal capacity (e.g., Harris et al. 2002; Kaliontzopoulou et al. 2011; Planas et al. 2013; Veith et al. 2004). Moreover, a connection of the north African coast and Sicily and Iberia may have existed during the Messinian Salinity Crisis (MSC) because of a decreased eustatic level and promoted dispersal of some taxa (Husemann et al. 2014). For ants, whose sexual castes are able to fly in most cases, these routes across the Mediterranean Sea seem likely. Northern Africa was suggested to be an important diversification centre during the tertiary glaciations, as southern European lineages of non-flying mammals, amphibians, reptiles, and insects such as butterflies, moths and dragonflies and damselflies are often nested within north African clades (Husemann et al. 2014). It is further an area of significant ant diversity within the Mediterranean context from which several lineages extending to southern Europe originated (Kass et al. 2022; Schifani et al. 2022). An ancestral south-western Mediterranean distribution (in northern Africa or the southernmost regions of Spain and Italy) of the western *C. sordidula* lineage therefore seems plausible on the basis of our reconstruction.

Interestingly, populations of *Crematogaster sordidula* belonging to the western and eastern lineages occur in close geographic proximity on the Italian Peninsula, yet remain genetically distant on the basis of the haplotype network, phylogenetic reconstruction, and population genomic clustering. More specifically, in the southern Peninsula, two samples group within the CEM clade in both the phylogenetic, haplotype, and admixture analyses and show no evidence of admixture with the remaining Italian samples. Gene flow is inferred only between these samples and the CEM clade, as well as regions of the Balkan Peninsula and the Aegean, but not between these samples and other Italian populations. This lends strong support for the presence of two sympatric but genetically isolated lineages. Biogeographic reconstructions further indicated two independent colonisation events of the Italian Peninsula—one from the southwestern Mediterranean and one from the eastern Mediterranean, potentially via dispersal across the Adriatic Sea (Wang et al. 2023). These results suggest that the eastern lineage likely represents a later arrival on the Italian Peninsula that now occurs in sympatry with, yet remains reproductively isolated from, the lineage historically established on the Peninsula. Expanded sampling of the entire Italian Peninsula is needed to reconstruct population history in this region.

4.3 | Lineage Diversification in the Eastern Mediterranean

The diversification of the eastern Mediterranean clades was likely highly influenced by the complex biogeographic history of this region. During the Pliocene–Pleistocene epochs, the Aegean region was affected by the repeated cycles of glaciation leading to sea level changes and alternating isolation or connection of the thousands of islands and islets in the Aegean Sea (Kornilios et al. 2019), temporarily isolating or reconnecting the fauna inhabiting those islands from neighbouring islands or the mainland. These paleoclimatic processes are assumed to have had a high impact on intraspecific diversification and distribution patterns of insects (Avtzis et al. 2021; Chatzimanolis et al. 2003; Trichas et al. 2020; Salata et al. 2020). For example, Avtzis et al. (2021) showed inter- and intraspecific divergences in beetles of the genus *Pygopleurus* on Aegean islands during the Pleistocene coinciding with the isolation of islands. Another study on *Dendarus* beetles recovered high haplotype diversity and population subdivision for Aegean populations, reflecting major palaeogeographical barriers due to regional variation in glaciation (Trichas et al. 2020).

Although our study cannot confirm the origin of *C. sordidula*, it seems plausible that the MRCA of the species arrived in the Mediterranean from northern Africa or entered the eastern Mediterranean via southern Türkiye, as our reconstructions revealed an MRCA for all eastern Mediterranean lineages occurring in Türkiye. This hypothesis also agrees with the early divergence of the EA clade in the Late Pliocene during the PPT, since an ancestral population would have already inhabited the southern part of Türkiye near the Aegean coast. This is supported by the intermediate position of the EA clade in the haplotype network, where it is found more closely related to samples from the AT clade than to the CEM populations (including Aegean region, Balkan Peninsula, Balkan mainland

and Adriatic coast). The East Aegean's unique haplotypes and distinct admixture group indicate early population isolation and subsequent stability. In the Early Pleistocene, the geography of the Aegean Sea was comparable with today's geography of the Aegean, with the mid-Aegean trench forming a natural barrier between the eastern and western Aegean region since the end of the MSC (around 5 Ma, Poulakakis et al. (2015)), pre-dating the occurrence of *C. sordidula* in that region. Our dating analysis suggests that during the end of the PPT, *C. sordidula* dispersed north-westwards over the Turkish Anatolian plateau across the Bosporus and Sea of Marmara into the southeastern Balkans.

4.4 | Colonisation and Diversification in the Central-Eastern Mediterranean Region

The Central-Eastern Mediterranean clade of *C. sordidula* dispersed over the Balkan Peninsula, Crete and Karpathos starting at ca. 2 Ma. The biogeographic analysis supports an initial range expansion from Türkiye to the Balkan Peninsula, followed by a range expansion to Crete and Karpathos, with subsequent population divergence on the islands. We found no evidence of admixture between populations on Crete/Karpathos or the central Aegean islands with populations of the EA or AT regions. In contrast, populations on the Balkan mainland show low levels of admixture with CEM populations and signals of gene flow between the regions were detected. This pattern supports a dispersal pathway first through the Balkan mainland followed by colonisation of the Balkan Peninsula, the central Aegean islands and the southernmost islands, Crete and Karpathos.

We estimated high genetic diversity and elevated heterozygosity on the Balkan mainland from our data, indicating a potential role of this area as refugium during Pleistocene glaciations. The Rhodope mountains in southern Bulgaria and the Thrace region of Greece may be candidates for plausible long-term refugial areas for *C. sordidula* in southeastern Europe (Hewitt 1999). Geological and glacial reconstructions indicate that, unlike the Alps and other high mountain systems, these mountain ranges did not host extensive ice sheets during the Pleistocene. Glaciation was restricted to the highest peaks, whereas the lower Rhodope range remained largely unglaciated (Hughes et al. 2006; Gachev et al. 2016). Our divergence time estimates for the CEM lineage of *C. sordidula* place the onset of diversification and regional structuring in the early to middle Pleistocene during a phase of intensifying glacial–interglacial cycles, associated with the onset of the MPT, which is consistent with a scenario of persistence in, and repeated expansion from, local refugia in the Rhodopes and adjacent areas. Evidence from other taxa supports the likelihood of a refugial role for the Rhodopes mountains; for example, the flowering plant *Haberlea rhodopensis* Friv., 1835 is inferred to be a tertiary relict occurring in a refugial area in southeastern Europe centred around the Rhodopes (Petrova et al. 2015). Chloroplast DNA data for the European hornbeam (*Carpinus betulus* Linnaeus, 1753) revealed the Pirin–Rhodope Mountains as one of the effective glacial refugia within the Balkans and as a source region for postglacial recolonisation (Postolache et al. 2017).

The high genetic distinctness and inferred Early–Mid Pleistocene history of *C. sordidula* populations in and around the Rhodopes seem, therefore, consistent with this area functioning as a long-term refugium and potential source for postglacial dispersal into the wider Balkan and Aegean regions. However, as for many taxa in the broader Balkan region, dense population sampling and genomic data, as well as robust time-calibrated phylogenies across co-distributed ant species, are still lacking. This limits our ability to resolve the precise timing, directionality, and demographic dynamics of refugial persistence and subsequent range expansions. Our interpretation of the Rhodope Mountains as a refugium for central-eastern lineages of *C. sordidula* must therefore remain a working hypothesis.

4.5 | The Role of Eastern Mediterranean Islands in the Diversification of *C. sordidula*

Our biogeographical reconstructions indicate an ancestral area for the CEM clade from the Balkan Peninsula to Crete and Karpathos, and thus a dispersal to Crete and Karpathos during the Pleistocene. Similarly, Kornilios et al. (2016) estimated a joint ancestral distribution on Crete and the Peloponnese for a clade of trapdoor spiders (genus *Cyrtocarenum*). Yet Crete has been isolated from the Peloponnese and the Cyclades since the MSC (around 5–6 Ma), and there were supposedly no land connections during the Pleistocene (Dermitzakis 1990). Potentially, the island's colonisation happened overseas during times of lower sea levels when the islands of Antikythera and Kythera could have acted as stepping stones, as proposed for the dispersal of a clade of tenebrionid beetles (*Dendarus* Latreille, 1829; Trichas et al. 2020). In our analysis, the most closely related *C. sordidula* lineage nested in the same haplotype cluster with samples from Crete is a sample from the neighbouring island, Karpathos. A second plausible dispersal route to Crete would be from the East Aegean mainland (i.e., Türkiye) via Rhodes, Karpathos, and Kasos, as also suggested for spiders (Kornilios et al. 2016) and beetles (Trichas et al. 2020). However, such a colonisation route was not supported by our admixture analysis, and no evidence of gene flow could be detected between the regions. Additional sampling from the south-east Aegean region would be needed to investigate this further.

Notably, we found four closely related unique haplotypes for populations on Crete and Karpathos with genetic divergence from the remaining CEM lineages, suggesting isolation from the remaining Aegean populations. This high genetic separation compared to other Aegean and Balkan Peninsula samples could be a result of a small effective population size and genetic drift typical for island populations (Furlan et al. 2012). Populations on Crete show the lowest genetic diversity compared to other populations within *C. sordidula*, potentially indicating a past genetic bottleneck on this island. It is possible that Crete has acted as a passive refugium during the Pleistocene glacial periods, containing relict populations not involved in post-glacial expansion (Médail and Diadema 2009). High-altitude areas of the Cretan mountains have been climatically stable during the Pleistocene climate oscillations, allowing taxa to survive climate oscillations in those areas.

In contrast to Crete, the smaller islands in the centre of the Aegean Sea, such as the Cyclades, have been connected, separated, and reconnected to the mainland several times during

the Tertiary glacial cycles (Hammoud et al. 2021), which had a profound impact on lineage diversification in this region. For example, diversification in flightless bush crickets was found to correlate with the massive changes in sea level during the Pliocene-Pleistocene glaciations in the Aegean region (Borissov et al. 2021). On the basis of our phylogenetic and biogeographic reconstructions, *C. sordidula* already occupied mainland Greece, the Peloponnese, and Crete before reaching the Cyclades islands. Multiple dispersal events to the islands during glacial cycles seem likely and have been shown for other flying insects such as tenebrionid beetles (Papadopoulou et al. 2009). Ongoing genetic exchange during the Pleistocene between mainland Greece and the Cyclades populations is probable, given the extensive sea level fluctuations in the Aegean area during this time (Papoulia 2016). This is supported by our admixture analysis, showing populations on the Cyclades admixed with populations from the Balkan Peninsula and mainland, Crete, the Adriatic coast, and Italy. The estimated gene flow between regions indicates strong connections between the populations in the Aegean region and the remaining populations in the CEM clade. Moreover, the haplotype analysis shows a shared haplotype occurring in the Cyclades, mainland Greece, and the Adriatic coast. Additionally, the admixture analysis revealed a relationship between populations on the Cyclades and Crete, forming a shared ancestral admixture cluster in the southern Aegean region.

5 | Conclusion

By applying a combination of UCE phylogenomics and population genomic approaches to investigate the population structure of the widespread Mediterranean ant *C. sordidula*, we reveal a complex history of dispersal and lineage diversification, shaped most likely by Quaternary climate and paleogeographic changes. Pleistocene glaciations possibly forced *C. sordidula* into glacial refugia on the Balkan mainland, the Iberian Peninsula, and Northern Africa. *Crematogaster sordidula* shows relatively deep genetic divergence between allopatric lineages from geographically distant regions in the Mediterranean, but also in two genetically distinct lineages occurring in sympatry on the Italian Peninsula. A taxonomic assessment was beyond the scope of our study, and further morphological investigations will be necessary to evaluate whether the species should be split into two or more taxa on the basis of the identified lineages. Ideally, samples from the entire distribution range of the species, including the Middle East and Central Asia, as well as the four described subspecies, would be included in such a study. Many widespread ant species in the Mediterranean may show similar patterns, and our study has provided a step towards uncovering some of the existing genetic diversity and understanding mechanisms of diversification. Such data are important to identify general patterns for the diversification of Mediterranean ant lineages, to refine future predictions for the effect of climate change on biodiversity, and define priority regions for conservation. Similar to *C. sordidula*, many Mediterranean taxa also occur in the Northern African and Middle Eastern part of the Mediterranean; however, future assessments of species and lineage diversity in these areas are necessary to fully understand insect evolution in the Mediterranean biodiversity hotspot.

Acknowledgements

We thank Isabelle Waurick for support during molecular data generation in the MfN laboratory facilities, and Michael Ohl and Rudolf Meier for administrative support of J.H.V.'s Master's thesis. Two anonymous reviewers provided valuable suggestions for improvement to our manuscript. Funding for this study was provided by the National Science Foundation (DEB-2019431 to B.B.B.), the Elsa-Neumann-Stipendium des Landes Berlin (to J.H.V.), and the Museum für Naturkunde Berlin. Open Access funding enabled and organized by Projekt DEAL.

Funding

This work was supported by the National Science Foundation, DEB-2019431, the Museum für Naturkunde Berlin, and the Elsa-Neumann Stipendium des Landes Berlin.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw sequence data have been deposited in the NCBI Sequencing Read Archive under BioProject PRJNA1321187, with accession numbers provided in Table S2. Assembled sequence contigs, data matrices, trees, and other relevant data files to reproduce this study are available under 10.5281/zenodo.16446113.

References

Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19: 1655–1664. https://doi.org/10.1101/gr.094052.109.

Allegrucci, G., M. Rampini, P. Gratton, V. Todisco, and V. Sbordoni. 2009. "Testing Phylogenetic Hypotheses for Reconstructing the Evolutionary History of *Dolichopoda* Cave Crickets in the Eastern Mediterranean." *Journal of Biogeography* 36: 1785–1797. https://doi.org/10.1111/j.1365-2699.2009.02130.x.

Allegrucci, G., E. Trucchi, and V. Sbordoni. 2011. "Tempo and Mode of Species Diversification in *Dolichopoda* Cave Crickets (Orthoptera, Rhaphidophoridae)." *Molecular Phylogenetics and Evolution* 60: 108–121. https://doi.org/10.1016/j.ympev.2011.04.002.

Andermann, T., A. M. Fernandes, U. Olsson, et al. 2018. "Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements." *Systematic Biology* 68: 32–46. https://doi.org/10.1093/sysbio/syy039.

Angelis, K., and M. Dos Reis. 2015. "The Impact of Ancestral Population Size and Incomplete Lineage Sorting on Bayesian Estimation of Species Divergence Times." *Current Zoology* 61: 874–885. https://doi.org/10.1093/czoolo/61.5.874.

Avtzis, D. N., V. Markoudi, V. Mizerakis, et al. 2021. "The Aegean Archipelago as Cradle: Divergence of the *Glaphyrid* Genus *Pygopleurus* and Phylogeography of *P. foina*." *Systematics and Biodiversity* 19: 346–358. https://doi.org/10.1080/14772000.2021.1884622.

Bankevich, A., S. Nurk, D. Antipov, et al. 2012. "Spades: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19: 455–477. https://doi.org/10.1089/cmb.2012.0021.

Benítez-Benítez, C., S. Martín-Bravo, C. S. Bjorå, et al. 2021. "Geographical vs. Ecological Diversification in Carex Section *Phacocystis* (Cyperaceae): Patterns Hidden Behind a Twisted Taxonomy." *Journal of Systematics and Evolution* 59: 642–667. https://doi.org/10.1111/jse.12731.

Blaimer, B. B. 2012. "Acrobat Ants Go Global—Origin, Evolution and Systematics of the Genus *Crematogaster* (Hymenoptera: Formicidae)." *Molecular Phylogenetics and Evolution* 65: 421–436. https://doi.org/10.1016/j.ympev.2012.06.028.

Blaimer, B. B., S. G. Brady, T. R. Schultz, M. W. Lloyd, B. L. Fisher, and P. S. Ward. 2015. "Phylogenomic Methods Outperform Traditional Multi-Locus Approaches in Resolving Deep Evolutionary History: A Case Study of Formicine Ants." *BMC Evolutionary Biology* 15: 271. https://doi.org/10.1186/s12862-015-0552-5.

Blatrix, R., C. Aubert, T. Decaens, C. Berquier, M.-C. Andrei-Ruiz, and C. Galkowski. 2020. "Contribution of a DNA Barcode to an Assessment of the Specificity of Ant Taxa (Hymenoptera: Formicidae) on Corsica." *European Journal of Entomology* 117: 420–429. https://doi.org/10.14411/eje.2020.046.

Blondel, J., J. Aronson, J. Y. Bodiou, G. Boeuf, and C. Fontaine. 2010. *The Mediterranean Region: Biological Diversity in Space and Time*. Oxford University Press.

Blumenstiel, B., K. Cibulskis, S. Fisher, et al. 2010. "Targeted Exon Sequencing by In-Solution Hybrid Selection." *Current Protocols in Human Genetics* 66: 18.14.11–18.14.24. https://doi.org/10.1002/0471142905.hg1804s66.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Borissov, S. B., A. Bobeva, B. Çıplak, and D. P. Chobanov. 2020. "Evolution of *Poecilimon jonicus* Group (Orthoptera: Tettigoniidae): A History Linked to the Aegean Neogene Paleogeography." *Organisms Diversity & Evolution* 20: 803–819. https://doi.org/10.1007/s13127-020-00466-9.

Borissov, S. B., K. G. Heller, B. Çıplak, and D. P. Chobanov. 2023. "Origin, Evolution and Systematics of the Genus *Poecilimon* (Orthoptera: Tettigoniidae)—An Outburst of Diversification in the Aegean Area." *Systematic Entomology* 48: 198–220. https://doi.org/10.1111/syen.12580.

Borissov, S. B., G. H. Hristov, and D. P. Chobanov. 2021. "Phylogeography of the *Poecilimon ampliatus* Species Group (Orthoptera: Tettigoniidae) in the Context of the Pleistocene Glacial Cycles and the Origin of the Only Thelytokous Parthenogenetic Phaneropterine Bush-Cricket." *Arthropod Systematics & Phylogeny* 79: 401–418. https://doi.org/10.3897/asp.79.e66319.

Borowiec, L. 2014. "Catalogue of Ants of Europe, the Mediterranean Basin and Adjacent Regions (Hymenoptera: Formicidae)."

Borowiec, L., and S. Salata. 2021. "Notes on Ants (Hymenoptera: Formicidae) From Western Greece." *Annals of the Upper Silesian Museum in Bytom, Entomology* 30: 1–25.

Borowiec, M. L. 2019. "Spruceup: Fast and Flexible Identification, Visualization, and Removal of Outliers From Large Multiple Sequence Alignments." *Journal of Open Source Software* 4: 1635. https://doi.org/10.21105/joss.01635.

Bouckaert, R. R., J. Heled, D. Kühnert, et al. 2014. "BEAST2: A Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 10: e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Branstetter, M. G., J. T. Longino, P. S. Ward, and B. C. Faircloth. 2017. "Enriching the Ant Tree of Life: Enhanced UCE Bait Set for Genome-Scale Phylogenetics of Ants and Other Hymenoptera." *Methods in Ecology and Evolution* 8: 768–776. https://doi.org/10.1016/j.cub.2017.03.027.

Bryant, D., R. R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and A. RoyChoudhury. 2012. "Inferring Species Trees Directly From Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis." *Molecular Biology and Evolution* 29: 1917–1932. https://doi.org/10.1093/molbev/mss086.

Cartes, J. E., F. Maynou, F. Sardà, D. Lloris, and S. Tudela. 2004. *The Mediterranean Deep-Sea Ecosystems: An Overview of Their Diversity, Structure, Functioning and Anthropogenic Impacts*. Centre for Mediterranean Cooperation.

Cassar, L. F., B. Nabholz, E. L. Reboud, et al. 2025. "Whole-Genome Data Shed Light on Speciation and Within-Species Differentiation of the *Papilio machaon* Complex Around the Mediterranean Basin." *Systematic Entomology* 50: 12675. https://doi.org/10.1111/syen.12675.

Chatzimanolis, S., A. Trichas, S. Giokas, and M. Mylonas. 2003. "Phylogenetic Analysis and Biogeography of Aegean Taxa of the Genus *Dendarus* (Coleoptera: Tenebrionidae)." *Insect Systematics & Evolution* 34: 295–312. https://doi.org/10.1163/187631203788964773.

Csősz, S., A. Alicata, F. Báthori, et al. 2025. "Integrative Taxonomy Reveals Inflated Biodiversity in the European *Temnothorax unifasciatus* Complex (Hymenoptera: Formicidae)." *Zoologica Scripta* 54: 33–49. https://doi.org/10.1111/zsc.12690.

Csősz, S., J. Heinze, and I. Mikó. 2015. "Taxonomic Synopsis of the Ponto-Mediterranean Ants of *Temnothorax nylanderi* Species-Group." *PLoS One* 10: e0140000. https://doi.org/10.1371/journal.pone.0140000.

Dapporto, L., M. Menchetti, V. Dincă, et al. 2024. "The Genetic Legacy of the Quaternary Ice Ages for West Palearctic Butterflies." *Science Advances* 10: eadm8596. https://doi.org/10.1126/sciadv.adm8596.

Dapporto, L., M. Menchetti, R. Vodă, et al. 2022. "The Atlas of Mitochondrial Genetic Diversity for Western Palaearctic Butterflies." *Global Ecology and Biogeography* 31: 2184–2190. 10.111/geb.13579.

de Lecocq Pletincx, N., X. Cerdà, K. Kiran, C. Karaman, A. Taheri, and S. Aron. 2024. "Ecological Diversification Preceded Geographical Expansion During the Evolutionary Radiation of *Cataglyphis* Desert Ants." *iScience* 27: 109852. https://doi.org/10.1016/j.isci.2024.109852.

Dermitzakis, M. D. 1990. "The Evolution of the Aegeis During the Late Cenozoic." *Geologica Balcanica* 20: 3–16.

Dos Reis, M., and Z. Yang. 2019. "Bayesian Molecular Clock Dating Using Genome-Scale Datasets." In *Evolutionary Genomics: Statistical and Computational Methods*, vol. 1910, 309–330. Humana Press. https://doi.org/10.1007/978-1-4939-9074-0_10.

Drummond, A. J., and A. Rambaut. 2007. "BEAST: Bayesian Evolutionary Analysis by Sampling Trees." *BMC Evolutionary Biology* 7: 214. https://doi.org/10.1186/1471-2148-7-214.

Erickson, K. L., A. Pentico, A. M. Quattrini, and C. S. McFadden. 2021. "New Approaches to Species Delimitation and Population Structure of Anthozoans: Two Case Studies of Octocorals Using Ultraconserved Elements and Exons." *Molecular Ecology Resources* 21, no. 1: 78–92. https://doi.org/10.1111/1755-0998.13241.

Faircloth, B. C. 2013. "Illumiprocessor: A Trimmomatic Wrapper for Parallel Adapter and Quality Trimming."

Faircloth, B. C. 2015. "Phyluce Is a Software Package for the Analysis of Conserved Genomic Loci." *Bioinformatics* 32: 786–788. https://doi.org/10.1093/bioinformatics/btv646.

Faircloth, B. C., and T. C. Glenn. 2012. "Not All Sequence Tags Are Created Equal: Designing and Validating Sequence Identification Tags Robust to Indels." *PLoS One* 7: e42543. https://doi.org/10.1371/journal.pone.0042543.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. "Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales." *Systematic Biology* 61: 717–726. https://doi.org/10.1093/sysbio/sys004.

Fattorini, S. 2002. "Biogeography of the Tenebrionid Beetles (Coleoptera, Tenebrionidae) on the Aegean Islands (Greece)." *Journal of Biogeography* 29: 49–67. https://doi.org/10.1046/j.1365-2699.2002.00656.x.

Feliner, G. N. 2011. "Southern European Glacial Refugia: A Tale of Tales." *Taxon* 60: 365–372. https://doi.org/10.1002/tax.602007.

Fisher, S., A. Barry, J. Abreu, et al. 2011. "A Scalable, Fully Automated Process for Construction of Sequence-Ready Human Exome Targeted Capture Libraries." *Genome Biology* 12: R1. https://doi.org/10.1186/gb-2011-12-1-r1.

Furlan, E., J. Stoklosa, J. Griffiths, et al. 2012. "Small Population Size and Extremely Low Levels of Genetic Diversity in Island Populations of the Platypus, *Ornithorhynchus anatinus*." *Ecology and Evolution* 2: 844–857. https://doi.org/10.1002/ece3.195.

Gachev, E., K. Stoyanov, and A. Gikov. 2016. "Small Glaciers on the Balkan Peninsula: State and Changes in the Last Several Years." *Quaternary International* 415: 33–54. https://doi.org/10.1016/j.quaint.2015.10.042.

Gantenbein, B., and C. R. Largiadèr. 2003. "The Phylogeographic Importance of the Strait of Gibraltar as a Gene Flow Barrier in Terrestrial Arthropods: A Case Study With the Scorpion *Buthus occitanus* as Model Organism." *Molecular Phylogenetics and Evolution* 28: 119–130. https://doi.org/10.1016/S1055-7903(03)00031-9.

Garrick, B. J. 2010. "Interval Analysis Versus Probabilistic Analysis." *Risk Analysis* 30. https://doi.org/10.1111/j.1539-6924.2010.01360.x.

Gat, J. R., and M. Magaritz. 1980. "Climatic Variations in the Eastern Mediterranean Sea Area." *Naturwissenschaften* 67: 80–87. https://doi.org/10.1007/BF01054689.

Goudet, J. 2005. "Hierfstat, a Package for R to Compute and Test Hierarchical F-Statistics." *Molecular Ecology Notes* 5: 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Guénard, B., M. D. Weiser, K. Gómez, N. Narula, and E. P. Economo. 2017. *The Global Ant Biodiversity Informatics (GABI) Database: Synthesizing Data on the Geographic Distribution of Ant Species (Hymenoptera: Formicidae)*. Okinawa Institute of Science and Technology Graduate University.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of Phyml 3.0." *Systematic Biology* 59: 307–321. https://doi.org/10.1093/sysbio/syq010.

Habel, J. C., M. Meyer, A. El Mousadik, and T. Schmitt. 2008. "Africa Goes Europe: The Complete Phylogeography of the Marbled White Butterfly Species Complex *Melanargia galathea m. Lachesis* (Lepidoptera: Satyridae)." *Organisms Diversity & Evolution* 8: 121–129. https://doi.org/10.1016/j.ode.2007.04.002.

Habel, J. C., T. Schmitt, and P. Müller. 2005. "The Fourth Paradigm Pattern of Post-Glacial Range Expansion of European Terrestrial Species: The Phylogeography of the Marbled White Butterfly (Satyrinae, Lepidoptera)." *Journal of Biogeography* 32: 1489–1497. https://doi.org/10.1111/j.1365-2699.2005.01273.x.

Hammoud, C., K. Kougioumoutzis, K. F. Rijsdijk, et al. 2021. "Past Connections With the Mainland Structure Patterns of Insular Species Richness in a Continental-Shelf Archipelago (Aegean Sea, Greece)." *Ecology and Evolution* 11: 5441–5458. https://doi.org/10.1002/ece3.7438.

Harris, D. J., S. Carranza, E. N. Arnold, C. Pinho, and N. Ferrand. 2002. "Complex Biogeographical Distribution of Genetic Variation Within Podarcis Wall Lizards Across the Strait of Gibraltar." *Journal of Biogeography* 29: 1257–1262. https://doi.org/10.1046/j.1365-2699.2002.00744.x.

He, D., A. Choi, K. Pipatsrisawat, A. Darwiche, and E. Eskin. 2010. "Optimal Algorithms for Haplotype Assembly From Whole-Genome Sequence Data." *Bioinformatics* 26: i183–i190. https://doi.org/10.1093/bioinformatics/btq215.

Herbert, T. D., K. T. Lawrence, A. Tzanova, L. C. Peterson, R. Caballero-Gill, and C. S. Kelly. 2023. "The Mid-Pleistocene Climate Transition." *Annual Review of Earth and Planetary Sciences* 51: 1–27. https://doi.org/10.1146/annurev-earth-032320-095936.

Hewitt, G. M. 1999. "Post-Glacial Re-Colonization of European Biota." *Biological Journal of the Linnean Society* 68: 87–112. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x.

Hewitt, G. M. 2001. "Speciation, Hybrid Zones and Phylogeography—Or Seeing Genes in Space and Time." *Molecular Ecology* 10: 537–549. https://doi.org/10.1046/j.1365-294x.2001.01202.x.

Hochkirch, A., M. Bilz, C. C. Ferreira, et al. 2023. "A Multi-Taxon Analysis of European Red Lists Reveals Major Threats to Biodiversity." *PLoS One* 18: e0293083. https://doi.org/10.1371/journal.pone.0293083.

Hölldobler, B., and E. O. Wilson. 1990. *The Ants*. Harvard University Press.

Hosner, P. A., B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball. 2016. "Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes)." *Molecular Biology and Evolution* 33: 1110–1125. https://doi.org/10.1093/molbev/msv347.

Hughes, P. D., J. C. Woodward, P. L. Gibbard, M. G. Macklin, M. A. Gilmour, and G. R. Smith. 2006. "Quaternary Glacial History of the Mediterranean Mountains." *Progress in Physical Geography* 30, no. 3: 334–364. https://doi.org/10.1191/0309133306pp481ra.

Husemann, M., T. Schmitt, F. E. Zachos, W. Ulrich, and J. C. Habel. 2014. "Palaearctic Biogeography Revisited: Evidence for the Existence of a North African Refugium for Western Palaearctic Biota." *Journal of Biogeography* 41: 81–94. https://doi.org/10.1111/jbi.12180.

Jablonski, D., D. Jandzik, P. Mikulíček, et al. 2016. "Contrasting Evolutionary Histories of the Legless Lizards Slow Worms (*Anguis*) Shaped by the Topography of the Balkan Peninsula." *BMC Evolutionary Biology* 16: 99. https://doi.org/10.1186/s12862-016-0669-1.

Kaliontzopoulou, A., C. PinhoI, D. J. Harris, and M. A. Carretero. 2011. "When Cryptic Diversity Blurs the Picture: A Cautionary Tale From Iberian and North African Podarcis Wall Lizards." *Biological Journal of the Linnean Society* 103: 779–800. https://doi.org/10.1111/j.1095-8312.2011.01703.x.

Kass, J. M., B. Guénard, K. L. Dudley, et al. 2022. "The Global Distribution of Known and Undiscovered Ant Biodiversity." *Science Advances* 8: eabp9908. https://doi.org/10.1126/sciadv.abp9908.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. "Mafft: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform." *Nucleic Acids Research* 30: 3059–3066. https://doi.org/10.1093/nar/gkf436.

Khélifi, N., M. Sarnthein, M. Frank, N. Andersen, and D. Garbe-Schönberg. 2014. "Late Pliocene Variations of the Mediterranean Outflow." *Marine Geology* 357: 182–194. https://doi.org/10.1594/PANGAEA.863957.

Knaus, B. J., and N. J. Grünwald. 2017. "vcfR: A Package to Manipulate and Visualize Variant Call Format Data in R." *Molecular Ecology Resources* 17: 44–53. https://doi.org/10.1111/1755-0998.12549.

Korba, J., V. Opatova, A. Calatayud-Mascarell, et al. 2022. "Systematics and Phylogeography of Western Mediterranean Tarantulas (Araneae: Theraphosidae)." *Zoological Journal of the Linnean Society* 196: 845–884. https://doi.org/10.1093/zoolinnean/zlac042.

Kornilios, P., E. Thanou, P. Kapli, A. Parmakelis, and M. Chatzaki. 2016. "Peeking Through the Trapdoor: Historical Biogeography of the Aegean Endemic Spider *Cyrtocarenum ausserer*, 1871 With an Estimation of mtDNA Substitution Rates for Mygalomorphae." *Molecular Phylogenetics and Evolution* 98: 300–313. https://doi.org/10.1016/j.ympev.2016.01.021.

Kornilios, P., E. Thanou, P. Lymberakis, Ç. Ilgaz, Y. Kumlutaş, and A. Leaché. 2019. "Genome-Wide Markers Untangle the Green-Lizard Radiation in the Aegean Sea and Support a Rare Biogeographical

Pattern." *Journal of Biogeography* 46: 552–567. https://doi.org/10.1111/jbi.13524.

Kühne, G., J. Kosuch, A. Hochkirch, and T. Schmitt. 2017. "Extra-Mediterranean Glacial Refugia in a Mediterranean Faunal Element: The Phylogeography of the Chalk-Hill Blue *Polyommatus coridon* (Lepidoptera, Lycaenidae)." *Scientific Reports* 7: 43533. https://doi.org/10.1038/srep43533.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. "Partitionfinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses." *Molecular Biology and Evolution* 34: 772–773. https://doi.org/10.1093/molbev/msw260.

Leigh, J. W., and D. Bryant. 2015. "PopART: Full-Feature Software for Haplotype Network Construction." *Methods in Ecology and Evolution* 6: 1110–1116. https://doi.org/10.1111/2041-210X.12410.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and Samtools." *Bioinformatics* 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lionello, P., P. Malanotte-Rizzoli, R. Boscolo, P. Alpert, V. Artale, and L. Li. 2006. *The Mediterranean Climate: An Overview of the Main Characteristics and Issues Mediterranean Climate Variability*, 1–26. Elsevier.

Lischer, H. E., L. Excoffier, and G. Heckel. 2014. "Ignoring Heterozygous Sites Biases Phylogenomic Estimates of Divergence Times: Implications for the Evolutionary History of Microtus Voles." *Molecular Biology and Evolution* 31: 817–831. https://doi.org/10.1093/molbev/mst271.

Lukić, D., T. Pinceel, F. Marrone, et al. 2021. "Pleistocene Allopatric Differentiation Followed by Recent Range Expansion Explains the Distribution and Molecular Diversity of Two Congeneric Crustacean Species in the Palaearctic." *Scientific Reports* 11: 22866. https://doi.org/10.1038/s41598-021-02164-8.

Martinez-Sañudo, I., M. A. Perotti, I. Carofano, G. Santoiemma, L. Marri, and L. Mazzon. 2024. "The Biogeographic Patterns of the Olive Fly and Its Primary Symbiont Candidatus *Erwinia dacicola* Across the Distribution Area of the Olive Tree." *Scientific Reports* 14: 22483. https://doi.org/10.1038/s41598-024-73055-x.

Mas-Peinado, P., D. Buckley, J. L. Ruiz, and M. García-París. 2018. "Recurrent Diversification Patterns and Taxonomic Complexity in Morphologically Conservative Ancient Lineages of *Pimelia* (Coleoptera: Tenebrionidae)." *Systematic Entomology* 43: 522–548.

Matschiner, M. 2022. "Species Tree Inference With SNP Data." In *112 Plant Comparative Genomics*, 23–44. Springer. https://doi.org/10.1007/978-1-0716-2429-6_2.

Matzke, N. J. 2013. *Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing*. University of California.

Matzke, N. J. 2014. "Model Selection in Historical Biogeography Reveals That Founder-Event Speciation Is a Crucial Process in Island Clades." *Systematic Biology* 63: 951–970. https://doi.org/10.1093/sysbio/syu056.

Médail, F., and K. Diadema. 2009. "Glacial Refugia Influence Plant Diversity Patterns in the Mediterranean Basin." *Journal of Biogeography* 36: 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x.

Menchetti, M., E. Schifani, F. García, et al. n. d. "DNA Barcode Reference Library for European Ants: A Roadmap for Phylogeography and Species Discovery." *Molecular Ecology Resources*.

Menchetti, M., G. Talavera, A. Cini, et al. 2021. "Two Ways To Be Endemic. Alps and Apennines Are Different Functional Refugia During Climatic Cycles." *Molecular Ecology* 30: 1297–1310. https://doi.org/10.1111/mec.15795.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. "Biodiversity Hotspots for Conservation Priorities." *Nature* 403: 853–858. https://doi.org/10.1038/35002501.

Naughton, I., N. D. Tsutsui, P. S. Ward, and D. A. Holway. 2024. "An Assemblage-Level Comparison of Genetic Diversity and Population Genetic Structure Between Island and Mainland Ant Populations." *Evolution* 78: 1685–1698. https://doi.org/10.1111/1755-0998.13241.

Nguyen, L. T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2014. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32: 268–274. https://doi.org/10.1093/molbev/msu300.

Noguerales, V., and J. Ortego. 2022. "Genomic Evidence of Speciation by Fusion in a Recent Radiation of Grasshoppers." *Evolution* 76: 2618–2633. https://doi.org/10.1111/evo.14508.

Page, A. J., B. Taylor, A. J. Delaney, et al. 2016. "SNP-Sites: Rapid Efficient Extraction of SNPs From Multi-Fasta Alignments." *Microbial Genomics* 2: e000056. https://doi.org/10.1099/mgen.0.000056.

Papadopoulou, A., I. Anastasiou, B. Keskin, and A. P. Vogler. 2009. "Comparative Phylogeography of Tenebrionid Beetles in the Aegean Archipelago: The Effect of Dispersal Ability and Habitat Preference." *Molecular Ecology* 18: 2503–2517. https://doi.org/10.1111/j.1365-294X.2009.04207.x.

Papoulia, C. 2016. "Late Pleistocene to Early Holocene Sea-Crossings in the Aegean: Direct, Indirect and Controversial Evidence." Géoarchéologie Des îles De Méditerranée—Geoarchaeology of the Mediterranean Islands.

Periáñez, R., and J. M. Abril. 2015. "Computational Fluid Dynamics Simulations of the Zanclean Catastrophic Flood of the Mediterranean (5.33 Ma)." *Palaeogeography, Palaeoclimatology, Palaeoecology* 424: 49–60. https://doi.org/10.1016/j.palaeo.2015.02.017.

Petrova, G., D. Moyankova, K. Nishii, et al. 2015. "The European Paleoendemic *Haberlea rhodopensis* (Gesneriaceae) has an Oligocene Origin and a Pleistocene Diversification and Occurs in a Long-Persisting Refugial Area in Southeastern Europe." *International Journal of Plant Sciences* 176, no. 6: 499–514.

Planas, E., C. Fernández-Montraveta, and C. Ribera. 2013. "Molecular Systematics of the Wolf Spider Genus *Lycosa* (Araneae: Lycosidae) in the Western Mediterranean Basin." *Molecular Phylogenetics and Evolution* 67: 414–428. https://doi.org/10.1016/j.ympev.2013.02.006.

Postolache, D., F. Popescu, L. Paule, et al. 2017. "Unique Postglacial Evolution of the Hornbeam (*Carpinus betulus* L.) in the Carpathians and the Balkan Peninsula Revealed by Chloroplast DNA." *Science of the Total Environment* 599–600: 1493–1502. https://doi.org/10.1016/j.scitotenv.2017.05.062.

Potts, A. J., T. A. Hedderson, and G. W. Grimm. 2014. "Constructing Phylogenies in the Presence of Intra-Individual Site Polymorphisms (2isps) With a Focus on the Nuclear Ribosomal Cistron." *Systematic Biology* 63: 1–16. https://doi.org/10.1093/sysbio/syt052.

Poulakakis, N., P. Kapli, P. Lymberakis, et al. 2015. "A Review of Phylogeographic Analyses of Animal Taxa From the Aegean and Surrounding Regions." *Journal of Zoological Systematics and Evolutionary Research* 53: 18–32. https://doi.org/10.1111/jzs.12071.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81: 559–575. https://doi.org/10.1086/519795.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. "Posterior Summarization in Bayesian Phylogenetics Using TRACER 1.7." *Systematic Biology* 67: 901–904. https://doi.org/10.1093/sysbio/syy032.

Ree, R. H., and I. Sanmartín. 2018. "Conceptual and Statistical Problems With the DEC+j Model of Founder-Event Speciation and Its Comparison With DEC via Model Selection." *Journal of Biogeography* 45: 741–749. https://doi.org/10.1111/jbi.13173.

Rohland, N., and D. Reich. 2012. "Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture." *Genome Research* 22: 939–946. https://doi.org/10.1101/gr.128124.111.

Salata, S., K. Kalarus, L. Borowiec, A. Trichas, and K. Kujawa. 2020. "How Estimated Ant Diversity Is Biased by the Sampling Method? A Case Study of Crete: A Mediterranean Biodiversity Hotspot." *Biodiversity and Conservation* 29: 3031–3050. https://doi.org/10.1007/s10531-020-02014-5.

Salvi, D., D. J. Harris, A. Kaliontzopoulou, M. A. Carretero, and C. Pinho. 2013. "Persistence Across Pleistocene Ice Ages in Mediterranean and Extra-Mediterranean Refugia: Phylogeographic Insights From the Common Wall Lizard." *BMC Evolutionary Biology* 13: 147. https://doi.org/10.1186/1471-2148-13-147.

Schär, S., M. Menchetti, E. Schifani, et al. 2020. "Integrative Biodiversity Inventory of Ants From a Sicilian Archipelago Reveals High Diversity on Young Volcanic Islands (Hymenoptera: Formicidae)." *Organisms Diversity & Evolution* 20: 405–416. https://doi.org/10.1007/s13127-020-00442-3.

Schifani, E., A. Alicata, M. Menchetti, et al. 2022. "Revisiting the Morphological Species Groups of West-Palearctic *Aphaenogaster* Ants (Hymenoptera: Formicidae) Under a Phylogenetic Perspective: Toward an Evolutionary Classification." *Arthropod Systematics & Phylogeny* 80: 627–648. https://doi.org/10.3897/asp.80.e84428.

Schifani, E., A. Scupola, and A. Alicata. 2020. "Morphology, Ecology and Biogeography of *Myrmecina sicula* André, 1882, Rediscovered After 140 Years (Hymenoptera, Formicidae)." *Biogeographia-The Journal of Integrative Biogeography* 35: 444. https://doi.org/10.21426/B635048444.

Seifert, B. 2018. *Ants of Northern and Central Europe*. Lutra Verlags-und Vertriebsgesellschaft.

Sherpa, S., D. Salvi, I. Silva-Rocha, et al. 2024. "Reconstructing the Complex Colonisation Histories of Lizards Across Mediterranean Archipelagos." *Journal of Biogeography* 51: 157–172. https://doi.org/10.1111/jbi.14739.

Stiller, J., N. G. Wilson, and G. W. Rouse. 2023. "Range-Wide Population Genomics of Common Seadragons Shows Secondary Contact Over a Former Barrier and Insights on Illegal Capture." *BMC Biology* 21: 129. https://doi.org/10.1186/s12915-023-01628-9.

Taberlet, P., and J. Bouvet. 1994. "Mitochondrial DNA Polymorphism, Phylogeography, and Conservation Genetics of the Brown Bear *Ursus arctos* in Europe." *Proceedings of the Royal Society of London, Series B: Biological Sciences* 255: 195–200. https://doi.org/10.1098/rspb.1994.0028.

Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. "Comparative Phylogeography and Postglacial Colonization Routes in Europe." *Molecular Ecology* 7: 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x.

Tagliacollo, V. A., and R. Lanfear. 2018. "Estimating Improved Partitioning Schemes for Ultraconserved Elements." *Molecular Biology and Evolution* 35: 1798–1811. https://doi.org/10.1093/molbev/msy069.

Talavera, G., and J. Castresana. 2007. "Improvement of Phylogenies After Removing Divergent and Ambiguously Aligned Blocks From Protein Sequence Alignments." *Systematic Biology* 56: 564–577. https://doi.org/10.1080/10635150701472164.

Tinaut, A., and F. Ruano. 2021. "Biogeography of Iberian Ants (Hymenoptera: Formicidae)." *Diversity* 13: 88. https://doi.org/10.3390/d13020088.

Trichas, A., M. Smirli, A. Papadopoulou, I. Anastasiou, B. Keskin, and N. Poulakakis. 2020. "Dispersal Versus Vicariance in the Aegean: Combining Molecular and Morphological Phylogenies of Eastern Mediterranean *Dendarus* (Coleoptera: Tenebrionidae) Sheds New Light on the Phylogeography of the Aegean Area." *Zoological Journal of the Linnean Society* 190: 824–843. https://doi.org/10.1093/zoolinnean/zlaa022.

Veith, M., C. Mayer, B. Samraoui, D. D. Barroso, and S. Bogaerts. 2004. "From Europe to Africa and Vice Versa: Evidence for Multiple Intercontinental Dispersal in Ribbed Salamanders (Genus *Pleurodeles*)." *Journal of Biogeography* 31: 159–171. https://doi.org/10.1111/j.1365-2699.2004.00957.x.

Wang, R., J. M. Kass, C. Galkowski, et al. 2023. "Geographic and Climatic Constraints on Bioregionalization of European Ants." *Journal of Biogeography* 50: 503–514. https://doi.org/10.1111/jbi.14546.

Ward, P. S., and B. B. Blaimer. 2022. "Taxonomy in the Phylogenomic Era: Species Boundaries and Phylogenetic Relationships Among North American Ants of the *Crematogaster scutellaris* Group (Formicidae: Hymenoptera)." *Zoological Journal of the Linnean Society* 194: 893–937. https://doi.org/10.1093/zoolinnean/zlab047.

Winker, K., T. C. Glenn, and B. C. Faircloth. 2018. "Ultraconserved Elements (UCEs) Illuminate the Population Genomics of a Recent, High-Latitude Avian Speciation Event." *PeerJ* 6: e5735. https://doi.org/10.7717/peerj.5735.

Yang, Z. 2007. "Paml 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24: 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yeo, D., A. Srivathsan, and R. Meier. 2020. "Longer Is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification." *Systematic Biology* 69: 999–1015. https://doi.org/10.1093/sysbio/syaa014.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. "Astral-Iii: Polynomial Time Species Tree Reconstruction From Partially Resolved Gene Trees." *BMC Bioinformatics* 19: 153. https://doi.org/10.1186/s12859-018-2129-y.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Phylogenetic relationships of *Crematogaster sordidula* in the Mediterranean region. (A) Concatenated ML cladogram (90%-0.98-spruceup) with support values (SH-aLRT and ufBS) and (B) concatenated ML tree (90%-0.98-spruceup) with branch lengths. (C) Coalescent tree (ASTRAL-90-98) with support values (LPP). **Figure S2:** Phylogenetic relationships of Crematogaster sordidula in the Mediterranean region. (A) Concatenated ML cladogram (80%-0.95-spruceup) with support values (SH-aLRT and ufBS) and (B) Concatenated ML tree (80%-0.95-spruceup) with branch lengths. (C) Concatenated ML cladogram (80%-0.97-spruceup) with support values (SH-aLRT and ufBS) and (D) concatenated ML tree (80%-0.97-spruceup) with branch lengths. **Figure S3:** Phylogenetic relationships of Crematogaster sordidula in the Mediterranean region. (A) Concatenated ML cladogram (80%-0.98-spruceup) with support values (SH-aLRT and ufBS) and (B) concatenated ML tree (80%-0.98-spruceup) with branch lengths. (C) Concatenated ML cladogram (80%-untrimmed_p) with support values (SH-aLRT and ufBS) and (D) concatenated ML tree (80%-untrimmed_p) with branch lengths. **Figure S4:** Phylogenetic relationships of Crematogaster sordidula in the Mediterranean region. (A) Concatenated ML cladogram (80%-untrimmed_np) with support values (SH-aLRT and ufBS) and (B)

concatenated ML tree (80%-untrimmed_np) with branch lengths. (C) Concatenated ML cladogram (90%-untrimmed_np) with support values (SH-aLRT and ufBS) and (D) concatenated ML tree (90%-untrimmed_np) with branch lengths. **Figure S5:** Phylogenetic relationships of Crematogaster sordidula in the Mediterranean region. (A) Concatenated ML cladogram (90%-0.95-spruceup) with support values (SH-aLRT and ufBS) and (B) concatenated ML tree (90%-0.95-spruceup) with branch lengths. (C) Concatenated ML cladogram (90%-0.97-spruceup) with support values (SH-aLRT and ufBS) and (D) concatenated ML tree (90%-0.97-spruceup) with branch lengths. **Figure S6:** Chronogram estimated using approximate likelihood in MCMCTree with the PAMLv4.9 package and the 90%-0.98-spruceup topology, showing results of the dating of the root of Crematogaster sordidula under consideration of the three outgroups and calibration points published by (Blaimer 2012a). 95% HPD intervals can be accessed in Table S8, referring to numbers beside nodes. The timescale indicates time steps of 2.5 Ma during the Late Pliocene, Pleistocene and Holocene. **Figure S7:** (A) Ancestral area estimation of Crematogaster sordidula in the Mediterranean region. The biogeographical analysis was performed under the DIVALIKE+j model in BioGeoBEARS using the maximum clade credibility tree resulting from SNAPP analysis. Probabilities for reconstruction are given as percentages in the text and are depicted as pies (complete data for range probabilities can be accessed in Table S8). Pie states are: M = Morocco; S = Spain + southern France; I = Italy; D = Adriatic Coast; P = Balkan Peninsula; A = Aegean region; B = Balkan Mainland; T = Türkiye; C = Crete and Karpathos; MT = Morocco-Türkiye (North African coast); PC = Balkan Peninsula-Crete. The timescale is shown in 0.5 Ma increments; major geological and climatic intervals are indicated below the timescale. The Pliocene–Pleistocene Transition (PPT), the Mid-Pleistocene Transition (MPT), and the Last Glacial Maximum (LGM) are marked. (B) Map of the Mediterranean region coloured according to the biogeographical analysis. **Figure S8:** Population genetic structure of Crematogaster sordidula inferred from ADMIXTURE (K = 3; Table S12) and PCA analyses and visualised in a geographic context. (A) Geographic distribution of ADMIXTURE ancestry proportions (K = 3) across the Mediterranean region. Pie charts at sampling localities represent individual ancestry components, coloured according to the five inferred clusters: C1 = red, C2 = yellow, C3 = blue. (B) Geographic distribution of admixed individuals under K = 3 across the Mediterranean region; (C) Principal component analysis (PCA) of all individuals on the basis of 2481 SNPs. The first two principal components are shown (PC1 = 49.77% and PC2 = 22.12%) representing 71.89% of the total variance. (D) ADMIXTURE results for the alternative model (K = 3), with individual ancestry proportions shown as vertical bars. **Figure S9:** Population genetic structure of Crematogaster sordidula inferred from ADMIXTURE (K = 4; Table S12) and PCA analyses and visualised in a geographic context. (A) Geographic distribution of ADMIXTURE ancestry proportions (K = 4) across the Mediterranean region. Pie charts at sampling localities represent individual ancestry components, coloured according to the five inferred clusters: C1 = red, C2 = yellow, C3 = blue. C4 = turquoise. (B) Geographic distribution of admixed individuals under K = 4 across the Mediterranean region; (C) principal component analysis (PCA) of all individuals on the basis of 2481 SNPs. The first two principal components are shown (PC1 = 49.77% and PC2 = 22.12%) representing 71.89% of the total variance. (D) ADMIXTURE results for the alternative model (K = 4), with individual ancestry proportions shown as vertical bars. **Figure S10:** Population genetic structure of Crematogaster sordidula inferred from ADMIXTURE (K = 6; Table S12) and PCA analyses and visualised in a geographic context. (A) Geographic distribution of ADMIXTURE ancestry proportions (K = 6) across the Mediterranean region. Pie charts at sampling localities represent individual ancestry components, coloured according to the five inferred clusters: C1 = red, C2 = yellow, C3 = blue. C4 = turquoise, C5 = grey, C6 = green. (B) Geographic distribution of admixed individuals under K = 6 across the Mediterranean region; (C) principal component analysis (PCA) of all individuals on the basis of 2481 SNPs. The first two principal components are shown (PC1 = 49.77% and PC2 = 22.12%) representing 71.89% of the total variance. (D) ADMIXTURE results for the alternative model (K = 6), with individual ancestry proportions shown as vertical bars. **Figure S11:** Population genetic structure of Crematogaster sordidula inferred from ADMIXTURE (K = 7; Table S12) and PCA analyses and visualised in a geographic context. (A) Geographic distribution of ADMIXTURE ancestry proportions (K = 7) across the Mediterranean region. Pie charts at sampling localities represent individual ancestry components, coloured according to the five inferred clusters: C1 = red, C2 = yellow, C3 = blue. C4 = turquoise, C5 = grey, C6 = green, C7 = purple. (B) Geographic distribution of admixed individuals under K = 5 across the Mediterranean region; (C) principal component analysis (PCA) of all individuals on the basis of 2481 SNPs. The first two principal components are shown (PC1 = 49.77% and PC2 = 22.12%) representing 71.89% of the total variance. (D) ADMIXTURE results for the alternative model (K = 7), with individual ancestry proportions shown as vertical bars. **Figure S12:** Pairwise genetic differentiation and spatial representation of Fst relationships within Crematogaster sordidula among 10 Mediterranean regions. (A) Geographic visualisation of pairwise genetic differentiation on the basis of Weir and Cockerham's Fst values among regions. Nodes represent regional centroids derived from sampling localities, and all pairwise regional comparisons are shown. Edge colour reflects Fst magnitude, with dark blue indicating low genetic differentiation and warmer colours (purple to red) indicating higher genetic differentiation. (B) Minimum spanning tree (MST) constructed from the pairwise Fst matrix, connecting all regions while minimising total genetic differentiation. Regions are labelled following the analysis with: A = Aegean region, B = Balkan Mainland, C = Crete and Karpathos, D = Adriatic Coast, I = Italy, M = Morocco, P = Balkan Peninsula, S = Spain + southern France, T = Türkiye, Y = Sicily. **Figure S13:** Pairwise genetic differentiation and spatial patterns of genetic connectivity among 11 regions for Mediterranean populations of Crematogaster sordidula: (A) heatmap of pairwise Fst values among regions, showing the full range of genetic differentiation. The colour scale spans the complete empirical range of observed Fst values (0–0.44), with blue indicating low Fst values (low genetic differentiation), yellow intermediate values and red high Fst (high genetic differentiation); (B) geographic representation of genetic connectivity inferred from pairwise Fst. Connections between regions are shown for low to moderate Fst values (0–0.2) to emphasise regions of higher inferred connectivity. Line colour transitions from dark blue, indicating lowest Fst values and highest connectivity, to purple, red and finally yellow to indicate progressively higher Fst values and lower inferred connectivity. Regions are labelled following the analysis with: A = Aegean region, B = Balkan Mainland, C = Crete and Karpathos, D = Adriatic Coast, I = Italy (WM clade), I2 = Italy (CEM clade), M = Morocco, P = Balkan Peninsula, S = Spain + southern France, T = Türkiye, Y = Sicily. **Figure S14:** Pairwise genetic differentiation and spatial representation of Fst relationships among 11 Mediterranean regions of Crematogaster sordidula. (A) Geographic visualisation of pairwise genetic differentiation on the basis of Weir and Cockerham's Fst values among regions. Nodes represent regional centroids derived from sampling localities, and all pairwise regional comparisons are shown. Edge colour reflects Fst magnitude, with dark blue indicating low genetic differentiation and warmer colours (purple to red) indicating higher genetic differentiation. (B) Minimum spanning tree (MST) constructed from the pairwise Fst matrix, connecting all regions while minimising total genetic differentiation. Regions are labelled following the analysis with: A = Aegean region, B = Balkan Mainland, C = Crete and Karpathos, D = Adriatic Coast, I = Italy (WM clade), I2 = Italy (CEM clade), M = Morocco, P = Balkan Peninsula, S = Spain + southern France, T = Türkiye, Y = Sicily. **Table S1:** Collection information for sequenced taxa. All sequences were newly generated for this study, except for one sample of *C. sordidula* already published in Blaimer et al. (2018) and one sample of *C. scutellaris* already published in Ward and Blaimer (2022). **Table S2:** UCE sequencing results and NCBI SRA accession numbers. **Table S3:** Summary of phylogenetic information content and AT/GC content for all analysed data subsets. Calculated with AMAS (Borowiec 2019). **Table S4:** Biogeographical regions of samples. Distribution matrices used as input for ancestral range reconstruction with BioGeoBEARS (SNP_BioGeoBEARS), calculation of haplotype networks in popART (COI_Haplo), and for calculating genetic diversity per region: M = Morocco, S = Spain+ southern France, I = Italy, D = Adriatic Coast, P = Balkan

Peninsula, A = Aegean region, B = Balkan Mainland, T = Türkiye, C = Crete and Karpathos, Y = Sicily. Distribution areas refer to Figures 3–6 and S7, S12–S15. **Table S5:** Information on secondary calibrations used to calibrate divergence dating analyses, including their phylogenetic placement. Node numbers refer to Figure S6. **Table S6:** Comparison of model scores and parameters for biogeographic analysis. Compared are 6 models for the chronogram inferred with SNAPP. The constrained model refers to the analysis with dispersal constraints employed, whereas the unconstrained model employed no dispersal constraints. **Table S7:** Haplotype network statistics for intNJ haplotype reconstruction; covering Tajima's D statistics, nucleotide diversity, segregating sites and parsimony informative sites. **Table S8:** Summary of dating and biogeography results. Shown are dating results estimated from the SNP and 90%-98 topologies. Ancestral area probability results for the SNP topology are shown only for the best-scoring model DIVALIKE, DIVALIKE + j and the ancestral ranges with the highest probabilities. M = Morocco, S = Spain + southern France, I = Italy, D = Adriatic Coast, P = Balkan Peninsula, A = Aegean region, B = Balkan Mainland, T = Türkiye, C = Crete and Karpathos. Dating and biogeographical results refer to Figures 3 and S7. **Table S9:** Haplotypes according to the intNJ haplotype network reconstruction; Haplotype clusters, samples and biogeographical region are indicated; M = Morocco, S = Spain+ southern France, I = Italy, D = Adriatic Coast, P = Balkan Peninsula, A = Aegean region, B = Balkan Mainland, T = Türkiye, C = Crete and Karpathos. Haplotypes refer to Figure 4. **Table S10:** Percentage of variance explained by principal components. **Table S11:** Cross-validation (CV) results for all *K* values tested in the ADMIXTURE analysis. Shown are results estimated from the 10-fold cross-validation test performed with the ADMIXTURE analysis on the basis of "one SNP per UCE loci" alignment. *K* values range from one to 10. **Table S12:** ADMIXTURE clustering per sample. Shown are results estimated from the ADMIXTURE analysis on the basis of "one SNP per UCE loci" alignment. *K* values range from two to 10, the first two cluster components per sample are shown and the status (pure or admixed) of the sample is indicated on the basis of a treshhold of 5% admixture between samples. ADMIXTURE clustering results refer to Figures 5 and Figures S8–S11. **Table S13:** Spatial distribution of genetic diversity. Shown are the results for quantification of genetic diversity using the statistical parameters Ho/He, F and pi per region. The full SNP matrix (200.862 SNPs) was used as input. Data were analysed using the vcfR package in R. The statistical significance of the tested parameters was tested using Kruskal–Wallis' test, Pearson's test and Dunn Kruskal–Wallis comparison in the same R package. M = Morocco, S = Spain + southern France, I = Italy, D = Adriatic Coast, P = Balkan Peninsula, A = Aegean region, B = Balkan Mainland, T = Türkiye, C = Crete and Karpathos, Y = Sicily. **Data S1:** zsc70049-sup-0015-Supinfo.zip.