Received: 18 January 2024 | Revised: 7 May 2024 | Accepted: 12 July 2024

DOI: 10.1111/zsc.12690

Zoologica Scripta | WILEY

ORIGINAL ARTICLE

Integrative taxonomy reveals inflated biodiversity in the European *Temnothorax unifasciatus* complex (Hymenoptera: Formicidae)

Sándor Csősz[1,2] | **Antonio Alicata**[3] | **Ferenc Báthori**[2] | **Christophe Galkowski**[4] | **Enrico Schifani**[5] | **Zalimkhan Yusupov**[6] | **Gábor Herczeg**[1,2] | **Matthew M. Prebus**[7,8,9]

[1]HUN-REN-ELTE-MTM Integrative Ecology Research Group, Budapest, Hungary

[2]Department of Systematic Zoology and Ecology, Institute of Biology, ELTE-Eötvös Loránd University, Budapest, Hungary

[3]Department of Biological, Geological and Environmental Sciences, University of Catania, Catania, Italy

[4]Independent Researcher, Bordeaux, France

[5]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

[6]Tembotov Institute of Ecology of Mountain Territories of RAS, Nalchik, Russia

[7]Social Insect Research Group, School of Life Sciences, Arizona State University, Tempe, Arizona, USA

[8]Department of Integrative Taxonomy of Insects, Institute of Biology, University of Hohenheim, Stuttgart, Germany

[9]KomBioTa – Center for Biodiversity and Integrative Taxonomy Research, University of Hohenheim and State Museum of Natural History, Stuttgart, Germany

Correspondence
Ferenc Báthori, Department of Systematic Zoology and Ecology, Institute of Biology, ELTE-Eötvös Loránd University, Pázmány Péter ave 1/C, Budapest 1117, Hungary.
Email: bathori.ferenc@ttk.elte.hu

Funding information
Directorate for Biological Sciences, Grant/Award Number: NSF CAREER DEB-1943626; National Research, Development, and innovation Fund (Hungary), Grant/Award Number: K 147781

Abstract

Temnothorax unifasciatus (Latreille, 1798) is a widely distributed pan-European species from the Iberian Peninsula to the Caucasus. This taxon's relatively high morphological variability prompts the taxonomists of earlier times and today to mention the morphologically different elements at specific or subspecific ranks. This paper aims to understand the population structure and genetic diversity within this lineage via integrative taxonomy, incorporating molecular phylogenetics, species delimitation analyses and multivariate analyses of continuous morphometric data from across the geographic range of the *T. unifasciatus* complex. Phylogenetic analyses yielded incongruent trees. The genealogical diversity index (gdi) and the confirmatory analyses on morphological data found only weak, ambiguous delimitations within the *unifasciatus* complex. The most highly supported scenario splits *T. brackoi* from the remaining *unifasciatus* complex with ambiguous support (gdi = 0.56). This scenario is supported by multivariate morphometry with 100% accuracy in classification success. Instead, our results suggest complex morphological and genetic population structuring within the broad range of *T. unifasciatus*. Therefore, we confirm the validity of two species,

T. brackoi Salata & Borowiec, 2019 and T. unifasciatus (Latreille, 1798), and propose five new junior synonymies, *T. cordieri* (Bondroit, 1918) syn. nov., *T. tauricus* (Ruzsky, 1902) syn. nov., *T. berlandi* (Bondroit, 1918) syn. nov., *T. unifasciatus staegeri* (Bondroit, 1918) syn. nov., *T. tuberum ciscaucasicus* (Arnol'di, 1977) syn. nov. with the latter. To achieve maximal taxonomic stability, we designated a lectotype for *Temnothorax unifasciatus* (Latreille, 1798).

KEYWORDS

molecular analysis, morphometrics, nomenclature, refugium, synonymy

1 | INTRODUCTION

Modern taxonomy aims to understand patterns in biodiversity by describing, classifying and naming living organisms (Mayr & Ashlock, 1991). Given that species are in most cases the basic units of biology, alpha-taxonomy (the recognition, description and naming of species) plays a substantial role in the discipline. For example, a stable classification scheme facilitates scientific communication among researchers and between researchers and the public; moreover, they enable conservation efforts, as well as ecological and comparative work, for example, comparative genomics, transcriptomics and phenomics. Contemporary biodiversity research, taxonomy and systematics rely on phenotypic and molecular data to understand where one taxon starts and another ends (Cicero et al., 2021; Deans et al., 2015; Padial et al., 2010). In decades past, researchers have used subjective and numeric morphological data and various molecular markers, including mitochondrial DNA (mtDNA), to explore relationships among species. However, since subjectively evaluated phenotypic traits and mitochondrial molecular markers often exhibit a significant level of uncertainty (Chan et al., 2022; Tobias et al., 2010) contemporary approaches increasingly use quantitative analyses of numeric morphological data and single-copy nuclear molecular markers, such as ultraconserved elements (UCEs; Faircloth et al., 2012). These methods are required to achieve more accurate estimates of genetic diversity within a particular lineage.

Geography- and habitat-driven variability is often demonstrated in ubiquitous lineages with broad geographic distributions (Butcher et al., 2009; Díaz-Tapia et al., 2020), such as the Western Palaearctic *Temnothorax unifasciatus* (Latreille, 1798) complex (which includes the nominal species and valid-deemed congeners, *T. brackoi* Salata & Borowiec, 2019, *T. cordieri* (Bondroit, 1918), *T. rougeti* (Bondroit, 1918) and *T. tauricus* (Ruzsky, 1902)). This widely distributed European lineage represents a crucial element of dry meadows or forest edges and can display high nest density on the ground across the continent from the Iberian Peninsula to the Caucasus. Because this taxon plays an important role in ecosystems (Giannetti et al., 2019; Seifert, 2018), *T. unifasciatus* – as we knew it in the last decades – has become a widely preferred laboratory model organism (Bordoni et al., 2019; Brunner et al., 2011; Foitzik et al., 2007; Stroeymeyt et al., 2007; Walter et al., 2011).

Despite the steady accumulation of knowledge gained in field and laboratory research, the taxonomy and nomenclature of the *T. unifasciatus* complex have been exceptionally turbulent in recent decades. The high number of junior synonyms and the long taxonomic history of the complex's name-giving species *T. unifasciatus* demonstrate how the authors in the 20th and 21st centuries zealously interpreted the observed morphological variations as racial or specific differences. As of the preparation of the manuscript (the end of 2023), eight junior synonyms and five valid subspecies are listed in the AntCat online database (Bolton, 2023).

The list has been rapidly changing, even today. For example, the former subspecies *T. unifasciatus cordieri* (Bondroit, 1918), a Corsican endemic, has been raised to species rank based on mtDNA evidence (Blatrix et al., 2020). Others, like *T. rougeti*, and *T. tauricus* (Ruzsky, 1902) formerly considered synonyms of *T. unifasciatus* (Casevitz-Weulersse & Galkowski, 2009) have been revived from synonymy (Borowiec & Salata, 2012), but the reasoning is vague. Furthermore, a new species, *T. brackoi* was recently described (Salata & Borowiec, 2019). These very recent taxonomic acts shed light on the fact that the taxonomy of this species complex suffers from great uncertainty. This instability in taxonomy and nomenclature hampers the transferability of findings between researchers and hinders fluent scientific discourse. Thus, the taxonomy and phylogeny of the lineage that we call the '*Temnothorax unifasciatus* complex' is urgently required.

Here, we aim to understand the population structure and genetic diversity within this lineage via an integrative taxonomic framework, incorporating inference of the molecular phylogeny, species delimitation analyses and

multivariate analyses of continuous morphometric data. We analysed an extensive set of continuous morphometric data via robust multivariate statistic procedures, NC-Clustering, combined with Partitioning Algorithm based on Recursive Thresholding (PART). The PART algorithm estimates the number of clusters in the data and assigns observations (i.e. specimens or samples) into partitions. We confirmed the morphological species hypothesis imposed by NC-PART via the ordinating principal component analysis (PCA). Finally, we tested the validity of the prior morphological species hypothesis imposed by the two exploratory processes via a Cross-Validated Linear Discriminant Analysis (CV-LDA). In parallel, we sampled 41 specimens from across the geographic range of the *unifasciatus* complex for targeted enrichment of UCEs, in an effort to identify the constituent lineages and their correspondence to morphological clusters.

This workflow allows us to recognize monophyletic lineages and better understand the intra- and interspecific variability of certain traits. Thereby we can avoid ill-defined species delimitations. Finally, we aimed to draw nomenclatural consequences, the final step in taxonomic work. Beyond phylogeny and taxonomic findings, we also discuss the association between morphological distance and phylogenetic (genealogical) divergence of species.

2 | MATERIALS AND METHODS

2.1 | Molecular phylogeny

2.1.1 | Molecular data generation

To generate the genetic data, we extracted DNA, prepared genomic libraries and performed targeted enrichment of ultraconserved elements (UCEs). We extracted DNA non-destructively from adult worker ants using a flame-sterilized size 2 stainless steel insect pin to pierce the cuticle of the head, mesosoma and gaster on the right side of the specimens, then used a DNeasy Blood and Tissue Kit (Qiagen, Inc., Hilden, Germany) following the manufacturer's protocols. We verified DNA extract concentration using a Qubit 3.0 Fluorometer (Invitrogen, Waltham, MA, U.S.A.). We input up to 50 ng of DNA, sheared to a target fragment size of 400–600 bp with a QSonica Q800R sonicator (Qsonica, Newtown, CT, U.S.A.), into a genomic DNA library preparation protocol (KAPA HyperPrep Kit, KAPA Biosystems, Wilmington, MA, U.S.A.). For targeted enrichment of UCEs, we followed the protocol of Faircloth et al. (2015) as modified by Branstetter et al. (2017) using a unique combination of iTru barcoding adapters (Glenn et al., 2019; BadDNA, Athens, GA, U.S.A.) for each sample. We performed enrichments on pooled, barcoded libraries using the catalogue version of the Hym 2.5Kv2A ant-specific RNA probes (Branstetter et al., 2017; Arbor Biosciences, Ann Arbor, MI, U.S.A.), which targets 2524 UCE loci in the Formicidae. We followed the library enrichment procedures for the probe kit, using custom adapter blockers instead of the standard blockers (Glenn et al., 2019; BadDNA, Athens, GA, U.S.A.), and left enriched DNA bound to the streptavidin beads during PCR, as described in Faircloth et al. (2015). Following post-enrichment PCR, we purified the resulting pools using SpeedBead magnetic carboxylate beads (Rohland & Reich, 2012; Sigma-Aldrich, St Louis, MO, U.S.A.) and adjusted their volume to 22 μL. We verified enrichment success and measured size-adjusted DNA concentrations of each pool with qPCR using a SYBR-FASTqPCR kit (Kapa Biosystems, Wilmington, MA, U.S.A.) and a Bio-Rad CFX96 RT-PCR thermal cycler (Bio-Rad Laboratories, Hercules, CA, U.S.A.), subsequently combining all pools into an equimolar final pool. We sequenced the final pool in one lane at Novogene (Sacramento, CA, U.S.A.) on Illumina HiSeq 150 cycle Paired-End Sequencing v4 runs (llumina, San Diego, CA, U.S.A.), along with other enriched libraries for unrelated projects.

2.1.2 | Molecular data processing

We processed the resulting raw reads with the PHYLUCE pipeline (Faircloth, 2016). Using these, we generated four datasets:

1. unifasciatus_full: the full 42 taxon dataset, including the outgroup taxon *T. congruus* Smith 1874. This dataset incorporates samples from previously published datasets (Prebus, 2017; Schifani et al., 2022).
2. unifasciatus_full_50best: the above dataset, filtered to 50 loci using the program SortaDate to accommodate divergence dating analysis;
3. unifasciatus_reduced: the above dataset with the outgroup removed, composed of allele-phased exonic data extracted from the newly generated data, focused on the *T. unifasciatus* complex;
4. unifasciatus_reduced_50best: the dataset above was filtered to 50 loci using SortaDate to accommodate species delimitation analyses;
5. unifasciatus_reduced_mtDNA: mitochondrial loci extracted from the newly generated data.

2.1.3 | Dataset construction

For the unifasciatus_full dataset, we followed the standard PHYLUCE protocol for processing UCEs in preparation

for phylogenomic analysis, aligning the monolithic unaligned FASTA file with the phyluce_align_seqcap_align command, using MAFFT (Katoh & Standley, 2013) as the aligner (--aligner mafft) and opting not to edge-trim the alignment (—no-trim). We trimmed the resulting alignments with the phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed command in PHYLUCE, which uses GBlocks ver. 0.91b (Castresana, 2000), using the following settings: b1 0.5, b2 0.5, b3 12, b4 7. After removing UCE locus information from taxon labels using the command phyluce_align_remove_locus_name_from_nexus_lines, we examined the alignment statistics using the command phyluce_align_get_align_summary_data and generated a dataset in which each locus contains a minimum of 85% of all taxa using the command phyluce_align_get_only_loci_with_min_taxa. After an initial run with IQTREE v2.1.2 (Minh et al., 2020), we used the 'Spruceup' pipeline (Borowiec, 2019a) to adjust the branch lengths for downstream analyses. We selected an overall lognormal cutoff of 95% for the entire dataset, except for one sample (Temnothorax_unifasciatus_FRA006_P1405); after inspecting the distance distribution plot for this sample, we made a manual cut-off of 0.4 due to its persistently long branch after initial trimming.

To accommodate divergence dating analyses, we generated the unifasciatus_full_50best dataset by first estimating individual locus trees with IQTREE 2, rooting the trees with phyxrr (Brown et al., 2017) and using SortaDate (Smith et al., 2018) to select the 50 best loci for divergence dating.

For the unifasciatus_reduced dataset, we first followed the allele-phasing protocol used in Andermann et al. (2019), implemented in PHYLUCE v1.6.6 for the 41 newly generated sequences. Then, we extracted exonic sequence data using the 'uce-to-protein' pipeline (Borowiec, 2019b) from the phased data: following allele phasing, we split the monolithic FASTA file into individual UCE loci using the command phyluce_assembly_explode_get_fastas_file. We removed locus identifiers from the taxon names in each locus file with the command-line tool *sed*, and used these files as input for the 'uce-to-protein' pipeline. We used RNA-Seq reads of *Temnothorax curvispinosus* (Mayr) (NCBI BioProject PRJNA450816) as a query reference for the pipeline. We aligned the amino acid translations for each locus using MAFFT-LINSI, and then used the aligned data as references to realign the extracted nucleotide data using the local version of TranslatorX (Abascal et al., 2010). We then trimmed the aligned loci using phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed using the settings detailed above. We inspected each locus alignment by eye using Aliview (Larsson, 2014) and discarded sequences that were poorly aligned and most likely contaminants or poorly assembled reads. We used AMAS (Borowiec, 2016) to calculate the number of taxa and matrix length for each locus, and retained loci that had ≥70% taxon presence and were ≥150 bp long. The resulting dataset contained 1062 allele-phased loci.

To accommodate the computational limits of downstream species delimitation analyses, we generated the unifasciatus_reduced_50best dataset, using posterior predictive checks (Bollback, 2002) as a filtering mechanism on the unifasciatus_reduced dataset. Because the complexity of the General Time Reversible (GTR) nucleotide substitution model usually accommodates the substitution rates and occurrence frequencies observed in large alignments, in most cases substitution model selection analyses select this model for phylogenomic datasets. However, the *adequacy* of this substitution model is often taken for granted. Posterior predictive checks address model adequacy by first analysing empirical data in a Bayesian framework, then using the posterior distributions of parameters to simulate data; finally, a test statistic is calculated for the empirical data and the simulated datasets. If the value calculated for the test statistic of the empirical data falls within the distribution of the simulated data, model adequacy is supported. We used a custom RevBayes script (Höhna et al., 2016; Prebus, 2021) to partition each protein-coding locus by codon position (positions 1 + 2 vs. 3) and to perform posterior predictive checks on each partition using GTR + G as a substitution model and multinomial likelihood as a test statistic, which is indicative of the overall model fit. We assessed model adequacy using a two-tailed t-test and size effect as criteria, then ranked each partition: first by *p*-value, then by size effect. We then selected the 50 best data partitions.

To investigate the mitochondrial phylogeny, we generated the unifasciatus_reduced_mtDNA dataset, composed of 13 mitochondrial genes. Because mitochondrial sequences often have large repetitive regions that can complicate assembly, we followed the advice of the NOVOplasty tutorial (Dierckxsens et al., 2017) and used trimmomatic 0.38 (Bolger et al., 2014) to trim adapter contamination and Illumina sequencing artefacts from the raw UCE reads, without additional quality trimming. For each set of raw reads, we used the associated adapter FASTA file generated by initial processing with illumiprocessor (ILLUMINACLIP:adapters.fasta:2:30:10), and omitted the LEADING, TRAILING and SLIDINGWINDOW options that are used by default in illumiprocessor. We used these adapter-cleaned reads as input for the program mitoFinder (Allio et al., 2020) using the mitogenome of *Vollenhovia emeryi* (NCBI BioProjectPRJNA278668) as a reference and MetaSpades (Nurk et al., 2017) for assembly. We used the longest contig output from mitoFinder as input for mitoBim (Hahn et al., 2013) to extend the

contig and to assess coverage depth, as viewed in Tablet (Milne et al., 2013). In cases where mitoFinder did not assemble the full mitogenome, we used the mitogenome of the closest related specimen, as determined from the UCE phylogenetic inference above, as a reference backbone for assembly with mitoBim. We annotated the final assemblies via the MITOs webserver (Bernt et al., 2013). We extracted DNA sequence data for mitochondrial protein-coding genes using the annotation results, aligned with MAFFT-LINSI and trimmed with gblocks using the settings for the unifasciatus_75t_2160t dataset.

2.1.4 | Phylogenetic inference

We analysed the unifasciatus_full dataset with maximum likelihood (ML) and summary coalescent (SC) approaches. Because the assumption that the evolutionary rates of sequence data are homogenous is often violated in empirical data (Buckley et al., 2001), we partitioned our UCE loci into sets of similarly evolving sites. To achieve this, we used the command phyluce_align_format_nexus_files_for_raxml which concatenates loci into a single alignment and generates a partition file for input into the SWSC-EN method (Tagliacollo & Lanfear, 2018). For the ML analysis, we used the resulting data blocks as input for partitioning in IQTREE ver. 2.1.2 (Minh et al., 2020), using the command -m MFP + MERGE. Because the combination of gamma and proportion of invariable sites (+I+G) has been demonstrated to result in anomalies in likelihood estimation (Sullivan & Swofford, 2001; Yang, 2006), we set the rate heterogeneity models to a subset that includes everything except the combination of gamma and proportion of invariable sites (-mrate E, I, G) and set the search algorithm to -rclusterf 10. We used the resulting partitioned dataset as input for maximum likelihood tree inference in IQTREE, using 1000 ultrafast bootstrap replicates (-B 1000). For the summary coalescent approach, we first generated trees for each individual locus using the partitions from the SWSC-EN method and the IQTREE settings above. Because nodes with low statistical support can introduce noise into SC analyses (Zhang et al., 2017), we first collapsed all nodes with bootstrap support <10 into polytomies using Newick utilities (Junier & Zdobnov, 2010). We then used these collapsed individual locus trees as input for the summary coalescent program ASTRAL-III (Zhang et al., 2017).

For the unifasciatus_reduced dataset, we repeated the above analyses (ML + SC), but diverged by partitioning the nucleotide data based on codon position (positions 1 + 2 vs. 3) instead of SWSC-EN partitions. For the SC analysis, we used a mapping file to force the two alleles from each sample to be monophyletic.

For the unifasciatus_reduced_mtDNA dataset, we partitioned each of the 13 mitochondrial genes by codon position (positions 1 + 2 vs. 3) and used these as input for ML analysis. Because the mitochondrial genome is inherited along the matriline and only very rarely undergoes recombination, we did not perform the SC analysis for this dataset.

The results of the unifasciatus_reduced and unifasciatus_reduced_mtDNA analyses were compared in a cophylogeny plot using the *cophylo* command from the R package *phytools* (Revell, 2012).

2.1.5 | Species delimitation

We used BPP v4.1.3 (Flouri et al., 2018) for species delimitation on the unifasciatus_reduced_50best dataset. Because species delimitation approaches using the multispecies coalescent, specifically BPP, have been criticized for treating any population bottleneck as a speciation event (Chambers & Hillis, 2020; Sukumaran & Knowles, 2017), we further evaluated our results using the genealogical diversity index (*gdi*; Jackson et al., 2017; Leaché et al., 2019). The *gdi* is an estimate of the overall degree of genetic divergence between two taxa, taking into account the combined effects of genetic isolation and gene flow (Jackson et al., 2017). This index is useful in the context of allopatric populations or populations organized on a geographic cline, as is the case in this study (see Figure 1). The *gdi* is continuous and is therefore problematic for binary approaches such as species delimitation, and therefore the results should be viewed in the context of additional evidence. However, it offers an estimate of where two populations are on the speciation continuum (Stankowski & Ravinet, 2021). BPP is a Bayesian method that uses the multispecies coalescent (MSC) model and the reversible-jump Markov chain Monte Carlo algorithm (rjMCMC) to calculate species delimitation probabilities. For all of the BPP analyses that follow, we used the topology inferred from the SC analysis of the 'unifasciatus_reduced' dataset as a guide tree, using the unifasciatus_reduced_50best sequence dataset as input. First, we estimated the species divergence time (*tau*) and population size (*theta*) parameters using the A00 analysis (parameter estimation with fixed delimitation scheme and fixed tree) on the SC tree, using several delimitation hypotheses delimiting the tree to between 2 and 11 species. For each run, we estimated *theta*, set alpha for *theta* and *tau* to 2 and adjusted the priors until observed shifts were lower than one order of magnitude, following Yang (2015). We set the number of MCMC generations to 800 K, sampled every two generations, with a 25% burn-in, and checked for convergence

FIGURE 1 Distribution map of *Temnothorax unifasciatus complex* lineages: *T. bracoi* (lilac diamonds), *T. cordieri* (yellow triangles), *T. tauricus* (blue diamonds), *T. unifasciatus_Caucasus* (red rectangles), West European *T. unifasciatus* s.str. (ocher circles).

with Tracer (Rambaut et al., 2018). Because the resulting parameters for each delimitation scheme were very similar, we used the mean of the resulting parameters as priors for *theta* and *tau* (2 0.008 and 2 0.0004, respectively) for an A10 analysis (species delimitation on a fixed tree). In this analysis, we estimated *theta* and set the MCMC to sample as above. We ran 10 independent analyses and then averaged the posterior probabilities for the presence of nodes across runs. We considered nodes with 95% posterior probability or greater as strongly supported. We then ran an A00 analysis using the resulting species delimitation scheme and calculated the *gdi* using the hierarchical approach described in Leaché et al. (2019), considering values of *gdi* <0.2 as unsupported (no delimitation), >0.7 as well supported (independently evolving species) and values in between as ambiguous.

2.1.6 | Divergence dating

We used MCMCTREE in PAML v4.9e (Yang, 1997, 2007) with the approximate likelihood calculation to decrease the overall analysis time (dos Reis & Yang, 2011), to estimate a chronogram using the unifasciatus_full_50best dataset as input. Because our dataset did not contain any fossil calibration points, we used a secondary calibration for the root, obtained from Prebus (2017). We set the clock to the independent rates model (clock = 2) and used the HKY85 substitution model (model = 4). We set the birth-death prior (BDparas) to 500 500 1 following Brown and Yang (2010), which suggests these priors for shallow phylogenies with intraspecific sampling (the original article used time units of 1 Ma: we adjusted the priors to reflect our units of 100 Ma). We set the gamma prior to the overall substitution rate for genes (rgene gamma) to 1 16 using the results from an initial clock analysis in baseml. Finally, we set the number of samples to 10,000 with a sampling frequency of 5000 and a burn-in of 5,000,000. We ran two analyses in parallel to confirm that the results were stable, and checked for convergence using the program Tracer (Rambaut et al., 2018). We summarized the results of the analysis using the R package MCMCtreeR (Puttick, 2019).

2.2 | Morphology

2.2.1 | Material examined

In this research, 18 continuous morphometric traits were measured on 405 workers belonging to 134 nest samples from West Europe to the Caucasus and Anatolian Turkey. The complete list of materials examined is given in Table S1. The material is deposited in the following institutions: HNHM (Hungarian Natural History Museum, Budapest, Hungary), MHNG (Muséum d'histoire naturelle, Genf, Switzerland), NHMB (Naturhistorisches Museum Basel, Switzerland), NHMW (Naturhistorisches Museum Wien, Austria), SCPC (Private collection of SC), RBINS (Royal Belgian Institute of Natural Sciences), SIZK (Schmaulhausen Institute of Zoology, Kyiv, Ukraine),

SMNG (Senckenberg Museum für Naturkunde Görlitz, Germany) and TUEC (Collection of Trakya University, Edirne, Turkie). Sampling sites of *Temnothorax unifasciatus* species complex are shown in Figure 1.

2.2.2 | Protocol for morphometric character recording

Morphometric characters is defined as in Csősz et al. (2015). All measurements were made in μm using a pin-holding stage, permitting rotations around *X*, *Y* and *Z* axes. An Olympus SZX12 stereomicroscope was used at a magnification of x192 for each character. In this work, morphometric measurements have been collected from workers only, because on the one hand, all relevant taxon types belong to worker caste, and on the other hand, gynes are collected relatively rarely, which makes statistical analysis of their morphometric data difficult. Morphometric data are provided in μm throughout the paper. All individuals were measured by SC. Definitions of morphometric characters are detailed in Table 1 and Figure 2a–e.

2.3 | Multivariate statistics – Arriving at final Morphospecies hypothesis

2.3.1 | Exploratory analyses via NC-PART clustering

The prior species hypothesis was generated based on workers via a combined application of NC clustering (Seifert et al., 2014) and Partitioning Based on Recursive Thresholding (PART) (Nilsen & Lingjaerde, 2013). The protocol was published by Csősz and Fisher (2016) which is now applied with the following specific setups: bootstrap iterations in PART were set to 'b = 1000', and the minimum size of clusters was set to 'minSize = 5' for both 'hclust' and 'kmeans'. The optimal number of clusters and the partitioning of samples are accepted as preliminary species hypotheses in every case when the two clustering methods, 'hclust' and 'kmeans' via PART, have arrived at the same conclusion.

2.3.2 | Hypothesis testing by confirmatory analyses

The validity of the prior morphospecies hypothesis was tested by LDA, in which scatterplots and a histogram illustrate the distribution of the individuals in a morpho space and Leave one out cross-validation LDA (LOOCV-LDA) returns a more conservative estimate of the validity of the particular model based on repeating leave-one-out

TABLE 1 Verbatim trait definitions for morphometric character recording.

Abbr.	Verbal definition of the trait	See:
CL	Maximum cephalic length in the median line; the head must be carefully tilted to the position with the true maximum. Excavations of hind vertex and/or clypeus, if any, reduce CL	(Figure 2a)
CS	Cephalic size; the arithmetic mean of CL and CWb	
CWb	Maximum width of the head capsule, measured posterior to the eyes	(Figure 2a)
EL	Maximum diameter of the compound eye	(Figure 2b)
FRS	Minimum distance between the frontal carinae	(Figure 2a).
ML	Mesosoma length from the caudalmost point of the propodeal lobe to the transition point between the anterior pronotal slope and anterior propodeal shield (preferentially measured in lateral view; if the transition point is not well defined, use dorsal view and take the centre of the dark-shaded borderline between pronotal slope and pronotal shield as the anterior reference point	(Figure 2b)
MW	Maximum mesosoma width; pronotal width in workers	(Figure 2e)
NOH	Maximum height of the petiolar node, measured in lateral view from the uppermost point of the petiolar node perpendicular to a reference line set from the petiolar spiracle to the imaginary midpoint of the transition between dorso-caudal slope and dorsal profile of the caudal cylinder of the petiole	(Figure 2b).
NOL	Length of the petiolar node, measured in lateral view from the petiolar spiracle to dorso-caudal corner of caudal cylinder. Do not erroneously take as a reference point the dorso-caudal corner of the helcium, which is sometimes visible	(Figure 2c)
PEH	Maximum petiole height. The chord of the ventral petiolar profile at the node level is the reference line perpendicular to which the maximum height of the petiole is measured	(Figure 2b)
PEL	Diagonal petiolar length measured from the base of the anteroventral subpetiolar process to the dorso-caudal corner of the caudal cylinder.	(Figure 2d)

(Continues)

TABLE 1 (Continued)

Abbr.	Verbal definition of the trait	See:
PEW	Maximum width of petiole	(Figure 2e)
POC	Postocular distance. Use a cross-scaled ocular micrometre and adjust the head to the measuring position of CL. Caudal measuring point: median occipital margin; frontal measuring point: median head at the level of the posterior eye margin	(Figure 2a)
PLST	distance between the most dorso-caudal point of the propodeal lobe and the centre of the propodeal stigma	(Figure 2c)
PPW	Maximum width of the postpetiole	(Figure 2e).
SL	Maximum straight line scape length excluding the articular condyle	(Figure 2a)
SPBA	The smallest distance between the lateral margins of the spines at their base. This should be measured in the dorsofrontal view since the wider parts of the ventral propodeum do not interfere with the measurement in this position. If the lateral margins of spines diverge continuously from the tip to the base, the smallest distance at the base is not defined. In this case, SPBA is measured at the level of the bottom of the interspinal meniscus	(Figure 2e)
SPST	distance between the centre of the propodeal stigma and the spine tip. The stigma centre refers to the midpoint defined by the outer cuticular ring but not to the centre of the real stigma opening that may be positioned eccentrically	(Figure 2b)
SPTI	the distance between spine tips in dorsal view; if spine tips are rounded or truncated, the centres of spine tips are taken as reference points	(Figure 2f)

process for every data point against the model and testing its predicted position.

The distribution map based on the studied samples was created using QGIS software version 3.10.6. (QGIS Development Team, 2020).

3 | RESULTS

3.1 | Phylogeny

3.1.1 | Molecular data processing

Following assembly and UCE extraction, the mean number of loci per sample was 2296, with a mean contig length of 1473 bases and a mean coverage score of 60.1x (see Table S2). Following alignment, trimming and filtering the full UCE dataset to loci with ≥85% taxon presence (including samples from previous studies), the unifasciatus_full dataset had 2152 loci, with a mean locus alignment length of 1412 bases. The concatenated matrix was 3 Mb in length, in which 126 kb were variable, 57 kb were parsimony informative, with 19.2% missing data. After phasing, extracting and filtering exonic loci from the reduced dataset, the unifasciatus_reduced dataset contained 1062 loci, with a mean locus alignment length of 593 bases. After filtering the unifasciatus_reduced dataset further using posterior predictive checks, the unifasciatus_reduced_50best dataset had a mean alignment length of 263 bases. The unifasciatus_reduced_50best dataset consisted of data blocks selected from UCEs that were all unique, that is, none of the codon position data blocks was extracted from the same UCE. The dataset was biased toward the codon positions 1 + 2 (78%, or 39 of 50 data blocks). We recovered 13 mitochondrial genes from all samples, except for one sample of *T. unifasicatus*, which was missing ATP8 (Table S2).

3.1.2 | Phylogenetic inference

The partitioning analysis of the unifasciatus_full dataset in IQTREE resulted in a 207-partition scheme. Model selection on the unifasciatus_full dataset resulted in the assignment of 27 unique substitution models. The resulting tree had strong overall support and generally concurred with the topology inferred from the ASTRAL analysis (see Figures S1 and S2).

The partitioning analysis of the unifasciatus_reduced dataset in IQTREE resulted in a 106-partition scheme, which was assigned to 36 unique substitution models. The resulting tree had strong overall support: *Temnothorax brackoi*, the *unifasciatus* complex and several of the morpho-geographical clusters had full bootstrap support, which often became weaker within clusters, especially along the backbone of *cordieri/unifascatus* (see Figure 3). The ASTRAL analysis mostly concurred with these general results, assigning a high posterior probability to the clades mentioned above, becoming weaker within *cordieri/unifasciatus*. The two analyses arrived at the same general topology, diverging only in regions where statistical support was weak in both (see Figure 3).

The partitioning analysis of the unifasciatus_reduced_mtDNA dataset in IQTREE resulted in a 9-partition scheme, each of which had a unique substitution model. The resulting tree had strong overall support, but the

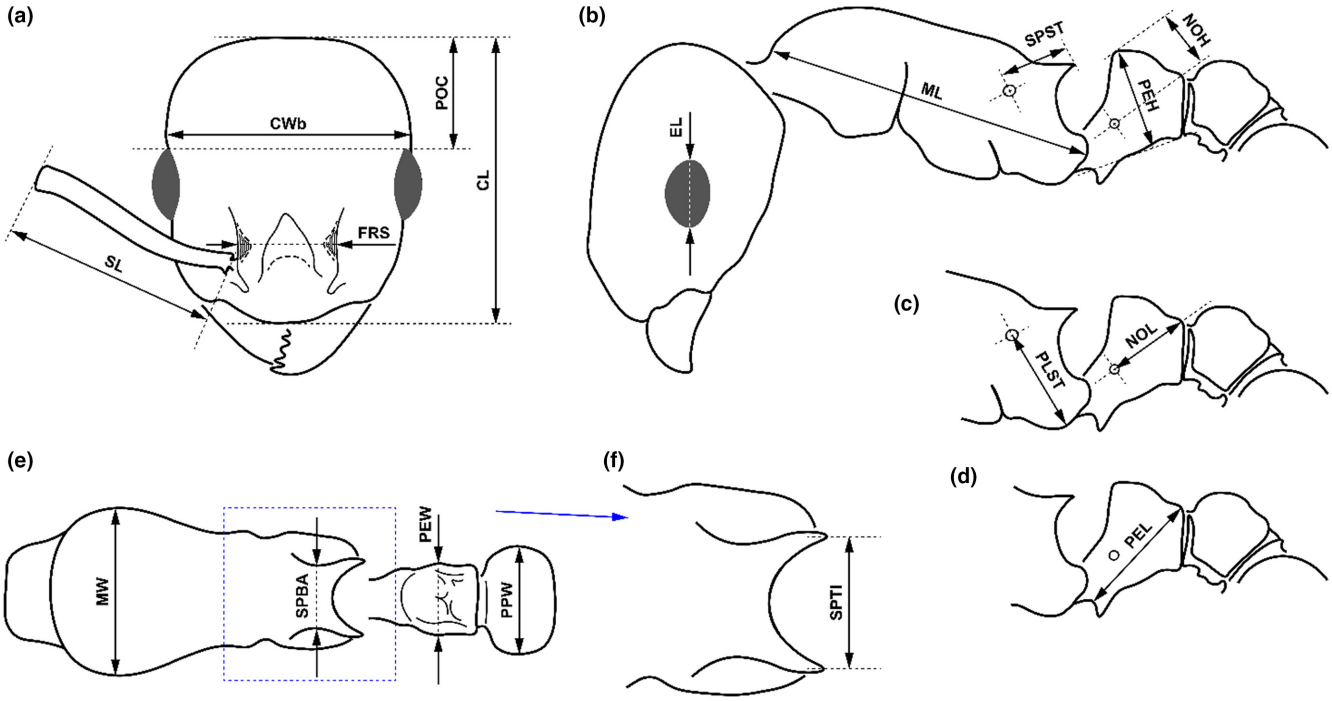

FIGURE 2 Illustrations for the measured morphometric characters.

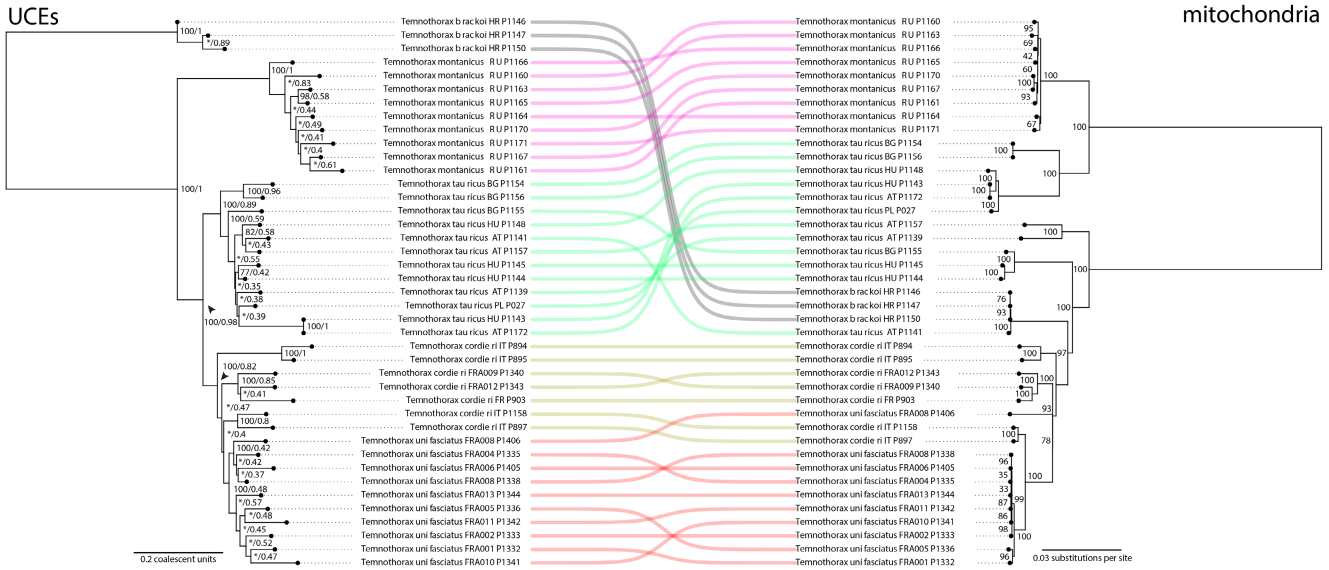

FIGURE 3 Results of phylogenomic analyses of the *Temnothorax unifasciatus* complex compared in a cophylogeny plot. Left: Phylogram resulting from analysis of the unifasciatus_reduced dataset with ASTRAL. Support values on each node are (ultrafast bootstrap from the IQTREE2 analysis)/(local posterior probability from the ASTRAL analysis); *indicates areas in which the results of the IQTREE2 analysis diverged in topology. Right: Phylogram resulting from analysis of the unifasciatus_reduced_mtDNA dataset with IQTREE2. Support values on each node are ultrafast bootstraps.

topology was radically different from that of the UCE analysis results. While some morpho-geographical clusters were highly supported, *tauricus* in particular was split up and placed into several clusters with high statistical support. Similarly to the UCE analyses, *cordieri* graded into *unifasciatus*.

3.1.3 | Species delimitation

The results of the initial BPP species delimitation analysis delimited the *unifasciatus* complex into 11 molecular operational taxonomic units (MOTUs). These roughly correspond to the morpho-geographical clusters found in the

morphometric analyses above but are more finely delimited (see Figure 4). In contrast, the *gdi* found only weak, ambiguous delimitations within the *unifasciatus* complex. The most highly supported scenario splits *T. brackoi* from the remaining *unifasciatus* complex with ambiguous support (*gdi* = 0.56; see Figure 4 and Table S3).

3.1.4 | Divergence dating

The two runs of the divergence dating analyses were identical, placing the emergence of the *unifasciatus* complex in the late Pliocene, ~3 Ma, with *unifasciatus* itself emerging ~2.1 Ma during the Gelasian age of the early Pleistocene, when ice sheets in the Northern Hemisphere began to grow. The deepest divergences between the morpho-geographical clusters span the Mid-Pleistocene Transition when glacial periodicity changed from low-amplitude to asymmetric high-amplitude cycles (Willeit et al., 2019; see Figure 5).

3.2 | Morphometrics

The two clustering methods 'hclust' and 'kmeans' of PART in combination with NC-clustering resulted in three (Figure 5) clusters. *Temnothorax brackoi* cluster is fully separated from the rest based on the two partitioning methods returned alternative assignments for 13 samples out of the total 123 *T. unifasciatus* complex samples. In the NC-clustering tree species deemed valid, *T. cordieri*, *T. tauricus*, *T. unifasciatus* and the Caucasian cluster were not recognized as separate clusters. The Central and East European *T. tauricus* was identified by partitioning methods in a slightly different way: 'part. hclust' enclosed *T. cordieri* samples from Corse and Sicily, and 'part.kmeans' added the same samples to the other cluster. The latter 'unifasciatus and Caucasian' subset encloses two remote populations, the West European *T. unifasciatus* and the Caucasian samples in both clustering methods.

The 2500 to 3000 km geographic gap between the West European *T. unifasciatus* and the Caucasian cluster hints at the chance that two cryptic lineages may exist in remote areas with deceptively similar morphology; hence we first set a five-cluster morphological species hypothesis (*T. brackoi*, *T. cordieri*, *T. tauricus*, *T. unifasciatus* and the Caucasian cluster) that was tested via confirmatory linear discriminant analysis (LDA) and LOOCV-LDA. The confirmatory process returned an overall accuracy = 0.86 (55 out of the 393 valid cases, Table 2, Figure 6). The LOOCV-LDA yielded a more

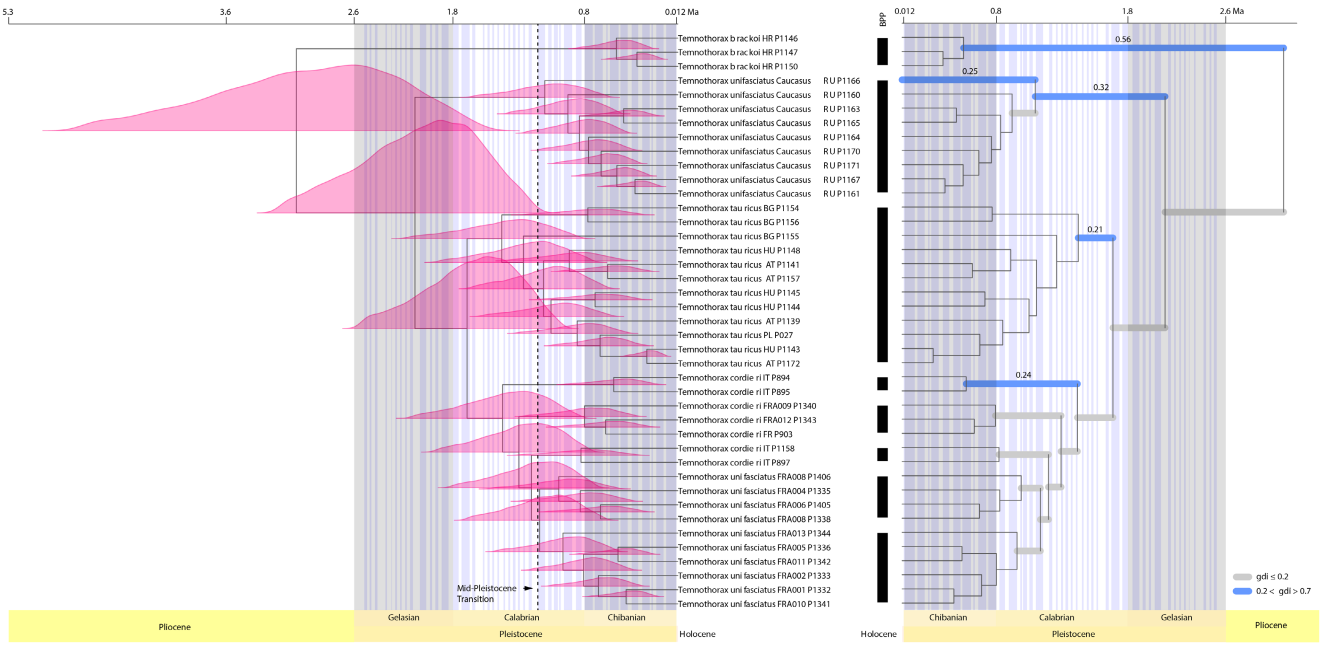

FIGURE 4 Divergence dating and species delimitation results plotted against glaciation cycles of the Pleistocene (Köhler & van de Wal, 2020). Distributions around nodes on the chronogram to the right indicate MCMC posterior density. Black bars between the two chronograms indicate the 11-species delimitation scheme suggested by the BPP analysis. Wide branches on the chronogram to the left indicate clades that were tested with the gdi calculation. Grey branches received no support for species status, while blue branches received ambiguous support; support values are given for clades with ambiguous support above the corresponding branches.

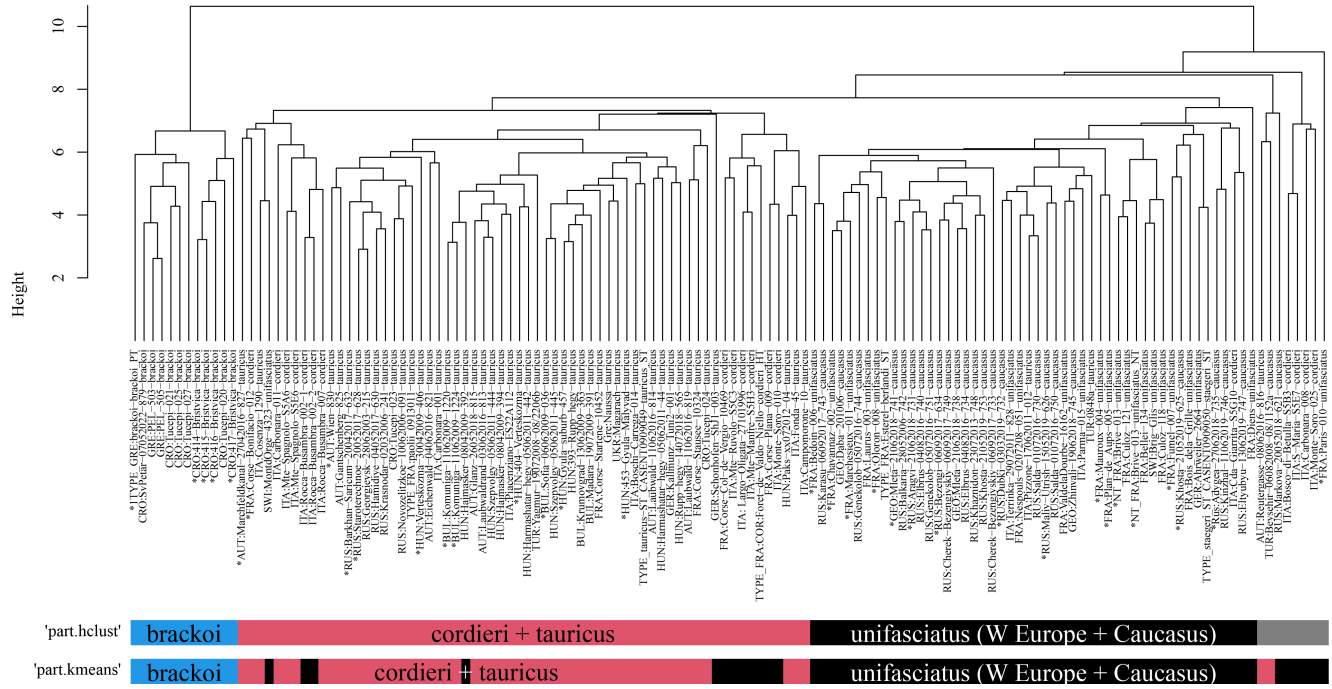

FIGURE 5 Dendrogram solution for morphometric data of *Temnothorax unifasciatus* complex species in NC Clustering using UPGMA distance method. The pattern is calculated from raw data, the labels represent nest samples. Bars represent 'kmeans', and 'hclust' partitioning results returned by function PART. Blue bars: *T. brackoi*, red bars: *T. cordieri* + *tauricus* cluster, black bars: *T. unifasciatus* s.str (from W-Europe) and the Caucasian *T. unifasciatus*. Grey bar in 'part.hclust' partitioning methods: Outliers.

TABLE 2 Classification table calculated via LDA using 18 traits on an individual level for the five-cluster hypothesis that is formed based on the currently valid-deemed European species.

	Correct%	Brackoi	unifasciatus_Caucasus	Cordieri	Tauricus	Unifasciatus s.str.
brackoi	100	31	0	0	0	0
unifasciatus_Caucasus	78.8	0	66	1	6	7
cordieri	75	0	0	44	11	0
tauricus	93.2	0	4	3	148	3
unifasciatus s.str.	59.1	0	9	1	10	49

pessimistic overall accuracy = 0.835, which is remote from the acceptable species criterion posed by the GAGE species concept (Seifert, 2020). The five-species hypothesis is rejected based on morphometry. The confirmatory tests of the three-species hypothesis returned by the NC-clustering yield overall accuracy = 0.931 (with 28 misclassified cases, Table 3, Figure 7). Only the *T. brackoi* separated from the rest with 100% certainty; hence we accept the two-species hypothesis: we have two species the Balkan endemism *T. brackoi* Salata & Borowiec, 2019 and the pan-European *T. unifasciatus* (Latreille, 1798) with four new junior synonyms, *T. berlandi* (Bondroit, 1918), *T. cordieri* (Bondroit, 1918), *T staegeri* (Bondroit, 1918) and *T. tauricus* (Ruzsky, 1902).

Updated taxonomy of Temnothorax unifasciatus complex

***T. brackoi* Salata & Borowiec, 2019**
***T. unifasciatus* (Latreille, 1798)**
Junior synonyms:
T. cordieri (Bondroit, 1918) **syn. nov.**
T. tauricus (Ruzsky, 1902) **syn. nov.**
T. berlandi (Bondroit, 1918) **syn. nov.**
T. unifasciatus staegeri (Bondroit, 1918) **syn. nov.**
T. tuberum ciscaucasicus (Arnol'di, 1977) **syn. nov.**
Infrasubspecific names:
Leptothorax tuberum unifasciatus paolii Santschi, 1923 [unavailable].

The species re-descriptions and differential diagnoses of the two valid species, *T. brackoi* Salata & Borowiec, 2019 and *T. unifasciatus* (Latreille, 1798), and the details of the neotype designation for the T. unifasciatus (Latreille, 1798) are provided as Appendix S1.

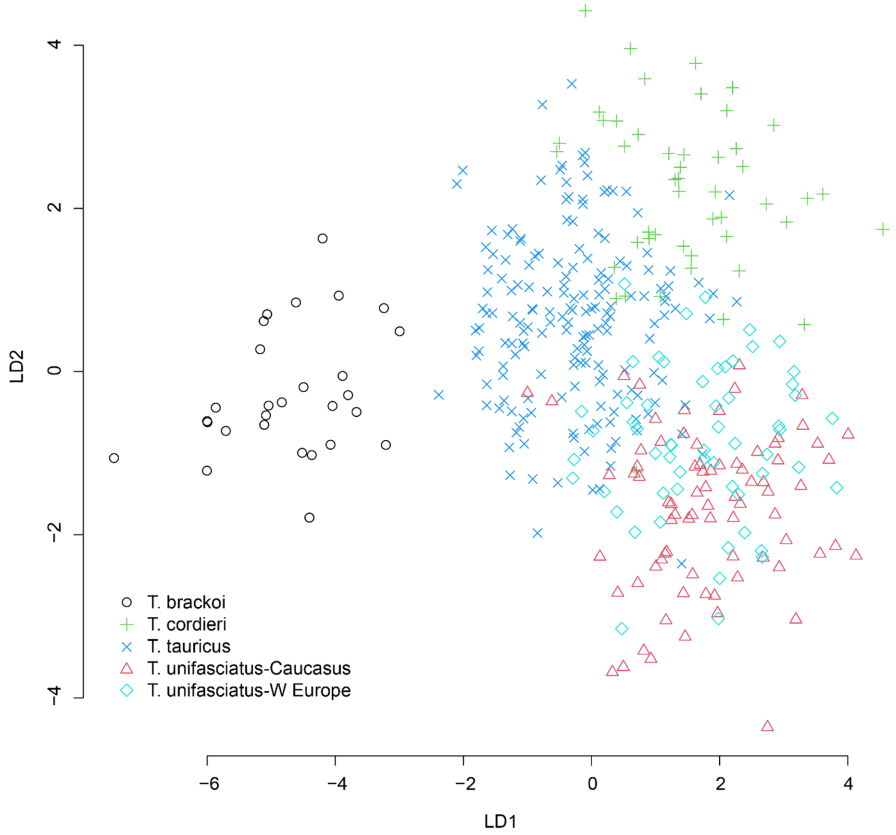

FIGURE 6 LDA plot for individual workers of *Temnothorax unifasciatus* complex species according to the currently valid classification: *T. brackoi (black circles)*, *T. cordieri* (green crosses), *T. tauricus* (blue crosses), *T. unifasciatus*-Caucasus (red triangles) West European *T. unifasciatus* s. str. (light blue diamonds).

TABLE 3 Classification table was calculated via LDA using 18 traits on an individual level for the three-cluster hypothesis returned by the exploratory NC-clustering. Headings: 'cor-tau': *T. cordieri* + *tauricus* cluster, 'uni-cau': Subset of the W-European *T. unifasciatus* s. str. and the Caucasus.

	Correct%	Brackoi	Cor-tau	Uni-cau
brackoi	100	32	2	0
cor-tau	91.3	0	199	9
uni-cau	94.2	0	17	146

4 | DISCUSSION

In this study, we initially recovered several candidate species from molecular and morphometric analysis of the *Temnothorax unifasciatus* complex, some of which correspond to described taxa. We achieved attaching Linnaean species names governed by the ICZN to the molecular lineages (Sluys, 2021; Steiner et al., 2009) via analysing a largely overlapping pool of samples with morphometric and molecular approaches, thereby making the two datasets compatible, and the results of the two methods comparable. Crucially, because we did not extract DNA from type specimens, we estimated which morphological cluster the morphometric data of the name-bearing type specimens fall into based on statistical thresholds.

In our integrative approach, we delineated independently evolving lineages, that is, species (de Queiroz, 2007) using the analysis of molecular and morphometric data. However, the first step in determining how many species (i.e. independently evolving metapopulation lineages) we have within this complex is selecting appropriate species recognition criteria from the nearly 30 existing examples (see de Queiroz, 1998, 2007; Sluys, 2021; Zachos, 2016; Zink, 1997). We adopted the GAGE species recognition criteria (Seifert, 2020) in the case of morphometric analyses because it relates to readily observable characteristics. It has been argued that in a taxonomic context, the preferred species recognition criteria should be associated with observable characteristics (Sluys, 2021); we chose this one because of its outstanding practical use. In the case of genetic data, we use an adaptation of the genealogical species recognition criteria (Baum & Shaw, 1995) suggested by Jackson et al. (2017), which sets thresholds for genealogical divergence when delimiting species in the presence of gene flow.

In an effort to evaluate the candidate species, we applied a molecular species delimitation method under the multispecies coalescent via BPP. Because this approach has been criticized as delimiting populations instead of species, and because our candidate species displayed geographical structure in phylogenetic analyses, we calculated the *gdi* as a further measure. We did not find any

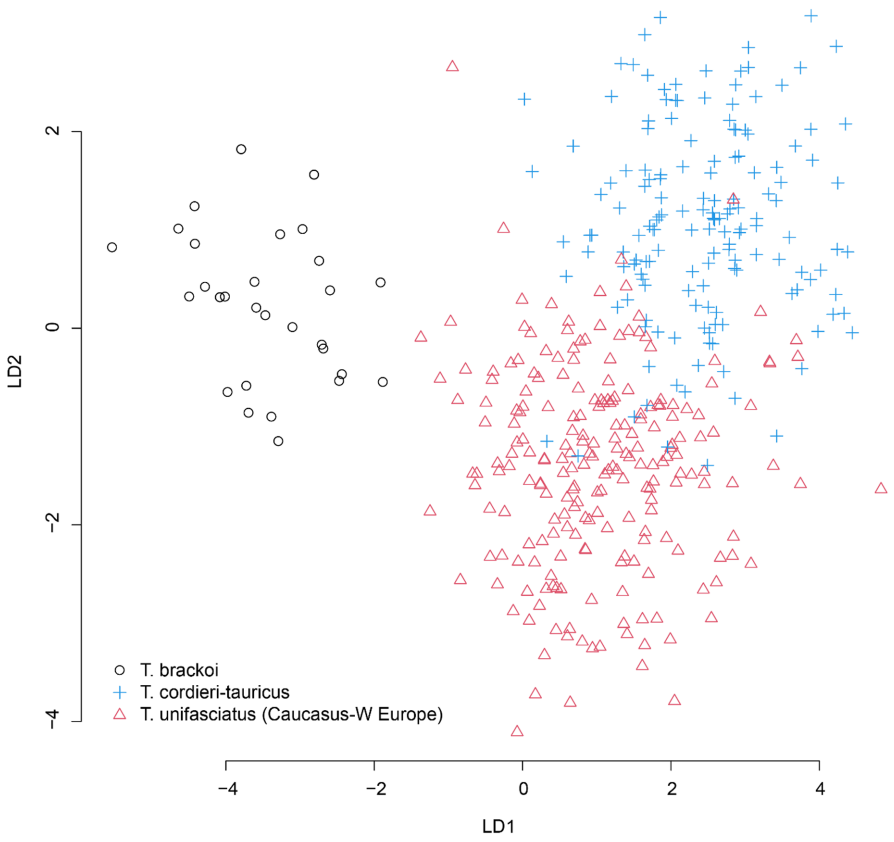

FIGURE 7 LDA plot for individual workers of *Temnothorax unifasciatus* complex species according to the pattern returned by the exploratory NC-clustering: *T. brackoi (black circles)*, *T. cordieri* + *T. tauricus* (blue crosses), *T. unifasciatus*-Caucasus and West Europe (red triangles).

definitive evidence of multiple species in the *unifasciatus* complex. Instead, our results suggest that the clustering of specimens recovered across analyses is indicative of morphological and genetic population structuring within the broad range of a single species.

From a morphometric point of view and following the guidelines of the GAGE species recognition criteria (Seifert, 2020), the received 86% overall classification success (resulting from the five-species hypothesis) using the entire character set does not meet species level criteria, in concordance with the *gdi* conclusion. The three-species hypothesis (i.e. *T. brackoi* and two *T. unifasciatus* metapopulations considered as separate classes) resulted in an overall accuracy of 93% (in which *T. brackoi* takes its share with 100%, and the two *T. unifasciatus* metapopulations show 91.3% and 94.2% respectively) is much more promising, but it is still below threshold. Based on these results, the species-level separation within *T. unifasciatus* populations/metapopulations cannot be supported morphometrically either.

In this paper, we propose formal junior synonymy for five taxa instead of proposing subspecies status for these geographically separate entities. We meticulously follow the integration by congruence protocol, which identifies boundaries of taxa with the intersection of evidence gained from two or more independent approaches instead of applying integration by cumulation protocol, which identifies boundaries with divergence in one or more not necessarily overlapping findings of an independent source of evidence (Padial et al., 2010). In contrast to the integration by cumulation (that may lead to lower confidence and overestimation of biodiversity) the integration by congruence protocol results in higher confidence, at the cost of possibly underestimating biodiversity. However, this rigorous approach ensures the accuracy and reliability of our proposed taxonomic changes.

In summary, the *Temnothorax unifasciatus* complex comprises two species, *T. brackoi* and *T. unifasciatus*. The former is a Balkan endemic, according to our current knowledge, with a narrow distributional area restricted to continental Greece and the Adriatic coast. *Temnothorax unifasciatus*, on the other hand, is a widely distributed species spreading from the Iberian Peninsula to the Caucasus. In its distribution area, it shows a strong population structure, which probably strengthened (and sometimes weakened) during the fluctuating glaciation cycles. Indeed, the divergence tree dates the deep divergences between populations of this species during the Pleistocene/early Holocene. This period is assumed to bolster speciation in many other species (Avise et al., 1998; Levsen et al., 2012), including many ant species (Goropashnaya et al., 2004; Leppänen et al., 2013). The fluctuating glacial expansions in the last few million years might have fragmented the connected populations and driven them into refugia, where they existed in isolation for a period. Then, in the interglacial periods, the

meeting populations either maintain their reproductive isolation or interbreed, resulting in the differences disappearing over time. The geographic distribution of the four infraspecific entities that we identified within *T. unifasciatus* suggests their origin to be linked to four southern European glacial refugia that played a pivotal biogeographical role during the Quaternary climatic fluctuations (Hewitt, 2000; Taberlet et al., 1998). These are presently important biodiversity hotspots in the West-Palearctic, which is reflected also in ant diversity patterns (Husemann et al., 2014; Wang et al., 2023). Iberia or southern France was likely the glacial refugium for *T. unifasciatus*, southern Italy for *T. cordieri*, the Balkans or Anatolia for *T. tauricus* and obviously Caucasus for the Caucasian entity. According to the divergence tree, *T. brackoi* diverged from the rest of the *T. unifasciatus* complex ca. 2.5 to 3 Mya and maintained its isolation in sympatry. The earliest dated divergence within *T. unifasciatus* is the Caucasian cluster that might have been isolated some 1.8 to 2 Mya, which is, according to the *gdi* and the insufficient discrimination in morphometric analyses, has not been enough to go through the speciation process. Then, the Central and East European *T. tauricus* cluster diverged in ca 1.5 Mya. The queue is closed by the West European *T. unifasciatus* s.str. and the South European *T. cordieri*, and according to the UCE analysis, *T. unifasciatus* s.str. can be considered the youngest cluster of all within this complex. In conclusion, the complex spatiotemporal population dynamics in the last few million years in Europe resulted in a considerable population structure in *T. unifasciatus*. However, the isolated entities have not gone through speciation.

The stability of classifications is considered a paramount need for users. It is particularly true if a ubiquitous species or a model species is considered. Our finding that the *Temnothorax unifasciatus* complex has been overestimated in its species-level diversity may inspire scientists to reconsider their results in taxonomy; the complicated spatiotemporal dynamics in the European fauna may result in unexpected intra- and interspecific variations. Therefore, integrative taxonomy might be desired to draw taxonomic conclusions.

ACKNOWLEDGEMENTS

This project has received funding from the HUN-REN Hungarian Research Network. This research was co-financed by the National Research, Development, and Innovation Fund (Hungary) under Grant No. K 147781 (On behalf of SC). MMP was supported by the Social Insect Research Group (SIRG) at the Arizona State University, and by a grant from the US National Science Foundation (NSF CAREER DEB-1943626).

ORCID

Sándor Csősz https://orcid.org/0000-0002-5422-5120
Antonio Alicata https://orcid.org/0000-0001-7762-2420
Ferenc Báthori https://orcid.org/0000-0001-5452-5257
Christophe Galkowski https://orcid.org/0000-0002-1001-1470
Enrico Schifani https://orcid.org/0000-0003-0684-6229
Zalimkhan Yusupov https://orcid.org/0000-0002-5149-9679
Gábor Herczeg https://orcid.org/0000-0003-0441-342X
Matthew M. Prebus https://orcid.org/0000-0001-8124-5939

REFERENCES

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, *38*(Suppl 2), W7–W13. https://doi.org/10.1093/nar/gkq291

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*, *20*(4), 892–905. https://doi.org/10.1111/1755-0998.13160

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B. C., & Antonelli, A. (2019). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Systematic Biology*, *68*(1), 32–46. https://doi.org/10.1093/sysbio/syy039

Avise, J. C., Walker, D., & Johns, G. C. (1998). Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1407), 1707–1712. https://doi.org/10.1098/rspb.1998.0492

Baum, D. A., & Shaw, K. L. (1995). Genealogical perspectives on the species problem. *Experimental and Molecular Approaches to Plant Biosystematics*, *53*(289–303), 123–124.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Blatrix, R., Aubert, C., Decaens, T., Berquier, C., Andrei-Ruiz, M. C., & Galkowski, C. (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

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

Bollback, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, *19*(7), 1171–1180. https://doi.org/10.1093/oxfordjournals.molbev.a004175

Bolton, B. (2023). *An online catalog of the ants of the world*. https://antcat.org

Bordoni, A., Matejkova, Z., Chimenti, L., Massai, L., Perito, B., Dapporto, L., & Turillazzi, S. (2019). Home economics in an oak gall: Behavioural and chemical immune strategies against

a fungal pathogen in *Temnothorax* ant nests. *The Science of Nature*, *106*, 1–8. https://doi.org/10.1007/s00114-019-1659-0

Borowiec, L., & Salata, S. (2012). Ants of Greece–checklist, comments and new faunistic data (hymenoptera: Formicidae). *Genus*, *23*(4), 461–563.

Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*, e1660. https://doi.org/10.7717/peerj.1660

Borowiec, M. L. (2019a). Spruceup: Fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. *Journal of Open Source Software*, *4*(42), 1635. https://doi.org/10.21105/joss.01635

Borowiec, M. L. (2019b). Convergent evolution of the army ant syndrome and congruence in big-data phylogenetics. *Systematic Biology*, *68*, 642–656. https://doi.org/10.1093/sysbio/syy088

Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (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*(6), 768–776. https://doi.org/10.1111/2041-210X.12742

Brown, J. W., Walker, J. F., & Smith, S. A. (2017). Phyx: Phylogenetic tools for unix. *Bioinformatics*, *33*(12), 1886–1888. https://doi.org/10.1093/bioinformatics/btx063

Brown, R. P., & Yang, Z. (2010). Bayesian dating of shallow phylogenies with a relaxed clock. *Systematic Biology*, *59*(2), 119–131. https://doi.org/10.1093/sysbio/syp082

Brunner, E., Kroiss, J., Trindl, A., & Heinze, J. (2011). Queen pheromones in *Temnothorax* ants: Control or honest signal? *BMC Evolutionary Biology*, *11*, 1–11. https://doi.org/10.1186/1471-2148-11-55

Buckley, T. R., Simon, C., & Chambers, G. K. (2001). Exploring among-site rate variation models in a maximum likelihood framework using empirical data: Effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Systematic Biology*, *50*(1), 67–86. https://doi.org/10.1080/10635150116786

Butcher, P. A., McDonald, M. W., & Bell, J. C. (2009). Congruence between environmental parameters, morphology and genetic structure in Australia's most widely distributed eucalypt, Eucalyptus camaldulensis. *Tree Genetics & Genomes*, *5*(1), 189–210. https://doi.org/10.1007/s11295-008-0169-6

Casevitz-Weulersse, J., & Galkowski, C. (2009). Up-to-date list of the ants of France (Hymenoptera, Formicidae). *Bulletin de la Société Entomologique de France*, *114*(4), 475–510.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chambers, E. A., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, *69*(1), 184–193. https://doi.org/10.1093/sysbio/syz042

Chan, K. O., Hertwig, S. T., Neokleous, D. N., Flury, J. M., & Brown, R. M. (2022). Widely used, short 16S rRNA mitochondrial gene fragments yield poor and erratic results in phylogenetic estimation and species delimitation of amphibians. *BMC Ecology and Evolution*, *22*(1), 1–9. https://doi.org/10.1186/s12862-022-01994-y

Cicero, C., Mason, N. A., Jiménez, R. A., Wait, D. R., Wang-Claypool, C. Y., & Bowie, R. C. (2021). Integrative taxonomy and geographic sampling underlie successful species delimitation. *The Auk*, *138*(2), ukab009. https://doi.org/10.1093/ornithology/ukab009

Csősz, S., & Fisher, B. L. (2016). Taxonomic revision of the Malagasy members of the *Nesomyrmex angulatus* species group using the automated morphological species delineation protocol NC-PART-clustering. *PeerJ*, *4*, e1796. https://doi.org/10.7717/peerj.1796

Csősz, S., Heinze, J., & Mikó, I. (2015). Taxonomic synopsis of the Ponto-Mediterranean ants of *Temnothorax nylanderi* species-group. *PLoS One*, *10*(11), e0140000. https://doi.org/10.1371/journal.pone.0140000

de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In D. J. Howard & S. H. Berlocher (Eds.), *Endless forms: Species and speciation* (pp. 57–75). Oxford University Press.

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

Deans, A. R., Lewis, S. E., Huala, E., Anzaldo, S. S., Ashburner, M., Balhoff, J. P., … Mabee, P. (2015). Finding our way through phenotypes. *PLoS Biology*, *13*(1), e1002033. https://doi.org/10.1371/journal.pbio.1002033

Díaz-Tapia, P., Ly, M., & Verbruggen, H. (2020). Extensive cryptic diversity in the widely distributed *Polysiphonia scopulorum* (Rhodomelaceae, Rhodophyta): Molecular species delimitation and morphometric analyses. *Molecular Phylogenetics and Evolution*, *152*, 106909. https://doi.org/10.1016/j.ympev.2020.106909

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

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

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among hymenoptera. *Molecular Ecology Resources*, *15*(3), 489–501. https://doi.org/10.1111/1755-0998.12328

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution*, *35*(10), 2585–2593. https://doi.org/10.1093/molbev/msy147

Foitzik, S., Sturm, H., Pusch, K., D'Ettorre, P., & Heinze, J. (2007). Nestmate recognition and intraspecific chemical and genetic variation in *Temnothorax* ants. *Animal Behaviour*, *73*(6), 999–1007. https://doi.org/10.1016/j.anbehav.2006.07.017

Giannetti, D., Castracani, C., Spotti, F. A., Mori, A., & Grasso, D. A. (2019). Gall-colonizing ants and their role as plant defenders: from 'bad job' to 'useful service'. *Insects*, *10*(11), 392. https://doi.org/10.3390/insects10110392

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L.,

Louha, S., Garcia-De Leon, F. J., Del Rio Portilla, M. A., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Belanger, M., … Faircloth, B. C. (2019). Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ*, *7*, e7755. https://doi.org/10.7717/peerj.7755

Goropashnaya, A. V., Fedorov, V. B., & Pamilo, P. (2004). Recent speciation in the *Formica rufa* group ants (Hymenoptera, Formicidae): Inference from mitochondrial DNA phylogeny. *Molecular Phylogenetics and Evolution*, *32*(1), 198–206. https://doi.org/10.1016/j.ympev.2003.11.016

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, *65*(4), 726–736. https://doi.org/10.1093/sysbio/syw021

Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W., & Habel, J. C. (2014). Palaearctic biogeography revisited: Evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography*, *41*(1), 81–94. https://doi.org/10.1111/jbi.12180

Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, *66*(5), 799–812. https://doi.org/10.1093/sysbio/syw117

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Köhler, P., & van de Wal, R. S. (2020). Interglacials of the Quaternary defined by northern hemispheric land ice distribution outside of Greenland. *Nature Communications*, *11*(1), 5124. https://doi.org/10.1038/s41467-020-18897-5

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The spectre of too many species. *Systematic Biology*, *68*(1), 168–181. https://doi.org/10.1093/sysbio/syy051

Leppänen, J., Vepsäläinen, K., Anthoni, H., & Savolainen, R. (2013). Comparative phylogeography of the ants *Myrmica ruginodis* and *Myrmica rubra*. *Journal of Biogeography*, *40*(3), 479–491. https://doi.org/10.1111/jbi.12026

Levsen, N. D., Tiffin, P., & Olson, M. S. (2012). Pleistocene speciation in the genus *Populus* (Salicaceae). *Systematic Biology*, *61*(3), 401–412. https://doi.org/10.1093/sysbio/syr120

Mayr, E., & Ashlock, P. D. (1991). *Principles of systematic zoology* (p. 475). McGraw-Hill, Inc.

Milne, I., Stephen, G., Bayer, M., Cock, P. J., Pritchard, L., Cardle, L., Shaw, P. D., & Marshall, D. (2013). Using tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, *14*(2), 193–202. https://doi.org/10.1093/bib/bbs012

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Nilsen, G., & Lingjaerde, O. C. (2013). *clusterGenomics: Identifying clusters in genomics data by recursive partitioning*. R package version 1.0. http://CRAN.R-project.org/package=clusterGenomics.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

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

Prebus, M. (2017). Insights into the evolution, biogeography and natural history of the acorn ants, genus *Temnothorax* Mayr (Hymenoptera: Formicidae). *BMC Evolutionary Biology*, *17*(1), 1–22. https://doi.org/10.1186/s12862-017-1095-8

Prebus, M. (2021). Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): An integrative approach. *Systematic Entomology*, *46*(2), 307–326. https://doi.org/10.1111/syen.12463

Puttick, M. N. (2019). MCMCtreeR: Functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics*, *35*(24), 5321–5322. https://doi.org/10.1093/bioinformatics/btz554

QGIS Development Team. (2020). QGIS Geographic Information System. (Version 3.10.6). Open Source Geospatial Foundation. http://qgis.org

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Reis, M. D., & Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Molecular Biology and Evolution*, *28*(7), 2161–2172. https://doi.org/10.1093/molbev/msr045

Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Salata, S., & Borowiec, L. (2019). Preliminary division of not socially parasitic Greek *Temnothorax* Mayr, 1861 (Hymenoptera, Formicidae) with a description of three new species. *ZooKeys*, *877*, 81–131. https://doi.org/10.3897/zookeys.877.36320

Schifani, E., Prebus, M. M., & Alicata, A. (2022). Integrating morphology with phylogenomics to describe four Island endemic species of *Temnothorax* from Sicily and Malta (Hymenoptera, Formicidae). *European Journal of Taxonomy*, *833*, 143–179. https://doi.org/10.5852/ejt.2022.833.1891

Seifert, B. (2018). *The ants of central and North Europe. Lutra Verlags – Und Vertriebsgesellschaft*. Tauer, Germany. 408 pp. [ISBN 9783936412079].

Seifert, B. (2020). The gene and gene expression (GAGE) species concept: An universal approach for all eukaryotic organisms. *Systematic Biology*, *69*(5), 1033–1038. https://doi.org/10.1093/sysbio/syaa032

Seifert, B., Ritz, M., & Csősz, S. (2014). Application of exploratory data analyses opens a new perspective in morphology-based alpha-taxonomy of eusocial organisms. *Myrmecological News*, *19*, 1–15.

Sluys, R. (2021). Attaching names to biological species: The use and value of type specimens in systematic zoology and natural history collections. *Biological Theory*, *16*(1), 49–61. https://doi.org/10.1007/s13752-020-00366-3

Smith, S. A., Brown, J. W., & Walker, J. F. (2018). So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One*, *13*(5), e0197433. https://doi.org/10.1371/journal.pone.0197433

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution*, *75*(6), 1256–1273. https://doi.org/10.1111/evo.14215

Steiner, F. M., Schlick-Steiner, B. C., & Seifert, B. (2009). Morphology-based taxonomy is essential to link molecular research to nomenclature. *Contributions to Natural History*, *12*, 1295–1315. https://doi.org/10.5169/seals-787027

Stroeymeyt, N., Brunner, E., & Heinze, J. (2007). "Selfish worker policing" controls reproduction in a *Temnothorax* ant. *Behavioral Ecology and Sociobiology*, *61*, 1449–1457. https://doi.org/10.1007/s00265-007-0377-3

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(7), 1607–1612. https://doi.org/10.1073/pnas.1607921114

Sullivan, J., & Swofford, D. L. (2001). Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Systematic Biology*, *50*(5), 723–729. https://doi.org/10.1080/106351501753328848

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*(4), 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Tagliacollo, V. A., & Lanfear, R. (2018). Estimating improved partitioning schemes for ultraconserved elements. *Molecular Biology and Evolution*, *35*(7), 1798–1811. https://doi.org/10.1093/molbev/msy069

Tobias, J. A., Seddon, N., Spottiswoode, C. N., Pilgrim, J. D., Fishpool, L. D., & Collar, N. J. (2010). Quantitative criteria for species delimitation. *Ibis*, *152*(4), 724–746. https://doi.org/10.1111/j.1474-919X.2010.01051.x

Walter, B., Brunner, E., & Heinze, J. (2011). Policing effectiveness depends on relatedness and group size. *The American Naturalist*, *177*(3), 368–376. https://doi.org/10.1086/658396

Wang, R., Kass, J. M., Galkowski, C., Garcia, F., Hamer, M. T., Radchenko, A., Salata, S. D., Schifani, E., Yusupov, Z. M., Economo, E. P., & Guénard, B. (2023). Geographic and climatic constraints on bioregionalization of European ants. *Journal of Biogeography*, *50*, 503–514. https://doi.org/10.1111/jbi.14546

Willeit, M., Ganopolski, A., Calov, R., & Brovkin, V. (2019). Mid-Pleistocene transition in glacial cycles explained by declining CO_2 and regolith removal. *Science Advances*, *5*(4), eaav7337. https://doi.org/10.1126/sciadv.aav7337

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, *13*(5), 555–556. https://doi.org/10.1093/bioinformatics/13.5.555

Yang, Z. (2006). *Computational molecular evolution* (p. 353). OUP Oxford.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/czoolo/61.5.854

Zachos, F. E. (2016). *Species concepts in biology* (Vol. *801*, p. 215). Springer. https://doi.org/10.1007/978-3-319-44966-1

Zhang, C., Sayyari, E., & Mirarab, S. (2017). ASTRAL-III: Increased scalability and impacts of contracting low support branches. In J. Meidanis & L. Nakhleh (Eds.), *Comparative genomics. RECOMB-CG 2017*. Lecture Notes in Computer Science (Vol. *10562*). Springer. https://doi.org/10.1007/978-3-319-67979-2_4

Zink, R. M. (1997). Species concepts. *Bulletin of the British Ornithologists' Club*, *117*(2), 97–109.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Csősz, S., Alicata, A., Báthori, F., Galkowski, C., Schifani, E., Yusupov, Z., Herczeg, G., & Prebus, M. M. (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