Ultraconserved yet informative for species delimitation: UCEs resolve long-standing systematic enigma in Central European bees

Morgan Gueuning[1], Juerg Frey[2], and Christophe Praz[1]

[1]Universite de Neuchatel Institut de Biologie
[2]Agroscope Standort Wadenswil

May 5, 2020

Abstract

Accurate and testable species delimitation hypotheses are essential for measuring, surveying and managing biodiversity. Today, taxonomists often rely on mitochondrial DNA barcoding to complement morphological species delimitations. Although COI barcoding has largely proven successful in assisting identifications for most animal taxa, there are nevertheless numerous cases where mitochondrial barcodes do not necessarily reflect the species history. For instance, what is regarded as one single species can be associated with two distinct DNA barcodes, which can point either to cryptic diversity or to deep within-species mitochondrial divergences with no reproductive isolation. In contrast, two or more species can share barcodes, for instance due to mitochondrial introgression. These intrinsic limitations of mitochondrial DNA barcoding can only be addressed with nuclear genomic markers, which are expensive, labour intensive, poorly repeatable, and often require high-quality DNA. To overcome these limitations, we examined the use of ultraconserved nuclear genetic elements (UCEs) as a quick and robust genomic approach to address such problematic cases of species delimitation. This genomic method was assessed using six different bee species complexes suspected to harbour cryptic diversity, mitochondrial introgression, or mitochondrial paraphyly. The sequencing of UCEs recovered between 686 and 1860 homologous nuclear loci and provided explicit species delimitation hypotheses in all investigated species complexes. These results provide strong evidence for the suitability of UCEs as a fast method for species delimitation even in recently diverged lineages. Furthermore, this study provided the first conclusive evidence for both mitochondrial introgression among distinct species, and mitochondrial paraphyly within a single bee species.

1. Introduction

Given the unprecedented levels of biodiversity losses currently observed (Barnosky et al., 2011; Pievani, 2014, uncovering cryptic diversity and providing testable species hypotheses is an urgent task for taxonomists, especially for the hyperdiverse group of the insects (Hallmann et al., 2017; Sánchez-Bayo & Wyckhuys, 2019; Seibold et al., 2019). Traditionally, species were described by examining variation in morphological traits (Padial, Miralles, De la Riva, & Vences, 2010). They were delimited in a way to minimize within-species variation and to maximise between-species variation in sets of variable characters. However, morphological taxonomy is often challenged by a lack of variation between taxa, or conversely by sexual or generational polymorphisms within species. Both of which will lead to an absence of a "morphological" gap between species and may result in substantial levels of cryptic diversity (Karanovic, Djurakic, & Eberhard, 2016). To complement morphology, DNA barcoding was introduced as a reliable, fast, and cheap identification method (Brunner, Fleming, & Frey, 2002; Hebert, Cywinska, Ball, & DeWaard, 2003), and has since been extensively used not only for specimen identification but also for species delimitation. For insects, the 5'-region of the cytochrome oxidase subunit I (COI) gene has quickly become the DNA barcode gold standard due to the fact that, for many species, it demonstrated only very limited intra-species variation (i.e. generally below

3%) yet distinct differentiation between species (e.g., Brunner et al., 2002; Hebert et al., 2003; Meyer & Paulay, 2005). In combination with morphological identification, COI-barcoding was shown to be a powerful tool for species delimitation in bees (Pauly, Noël, Sonet, Notton, & Boevé, 2019; Praz, Müller, & Genoud, 2019; Schmidt, Schmid-Egger, Morinière, Haszprunar, & Hebert, 2015).

There are, nevertheless, numerous examples where COI-barcoding leads to an erroneous signal. A number of possible reasons for such problematic barcoding results have recently emerged. For example, a growing body of literature is reporting that mitochondrial inheritance is more complicated than initially thought, with rare cases of paternal leakage, heteroplasmy or recombination (Ladoukakis & Zouros, 2017; White, Wolff, Pierson, & Gemmell, 2008). Furthermore, mitochondrial genomes can be subject to evolutionary forces acting solely at the organelle level [e.g. mitochondrial introgression, *Wolbachia* infection or sex-biased asymmetries; (Toews & Brelsford, 2012)]. Although these events are generally considered rare (but see Klopfstein, Kropf, & Baur, 2016; Neumeyer, Baur, Guex, & Praz, 2014; Nichols, Jordan, Jamie, Cant, & Hoffman, 2012), they can considerably skew phylogenies or biodiversity estimates (Andriollo, Naciri, & Ruedi, 2015; Hinojosa et al., 2019; Mutanen et al., 2016). Consequently, species delimitation should rely on multiple sources of information (Carstens, Pelletier, Reid, & Satler, 2013) and for molecular markers, species delimitation should use genes of both mitochondrial and nuclear origin (Dupuis, Roe, & Sperling, 2012).

In contrast to the uncommon suitability of COI as a species marker, the quest for similarly well-suited, universal and robust nuclear markers was so far unsuccessful. Several types of nuclear markers have been explored, but current candidates are all associated with serious drawbacks. For instance, single-copy nuclear genes (i.e. elongation factor 1 alpha [EF-1a] or 28S) or multicopy ribosomal DNA markers (i.e. internal transcribed spacer [ITS]) were explored (Leneveu, Chichvarkhin, & Wahlberg, 2009; Martinet et al., 2018; Soltani, Bénon, Alvarez, & Praz, 2017; Williams, Lelej, & Thaochan, 2019). However, the usefulness of these nuclear markers is often limited by the lack of phylogenetic resolution (Dellicour & Flot, 2018), and in insects, by within-genome variation of the multi-copy ribosomal genes (e.g. ITS), which is a major impediment to the sequencing workflow. For increased resolution, some studies have used population genetic markers such as microsatellites. Although microsatellites provide ample resolution for species delimitation (McKendrick et al., 2017), one major limitation is that loci are clade-specific and therefore require clade specific primers that have to be designed base on available genomic information. Alternative approaches that are slightly more universal and provide equal or higher resolution include genomic-reduction techniques such as RAD- or ddRAD sequencing (Lemopoulos et al., 2019). These methods are very powerful and can provide high-resolution, intraspecific information on population dynamics. However, as for microsatellites, RAD-seq is hampered by cost, workload and/or amount of high-quality DNA required. More importantly, datasets obtained from different studies and/or taxa are hardly joinable due to the lack of repeatability. This last limitation of RAD-sequencing is a severe drawback for species delimitation, since taxonomic work essentially builds upon previous hypotheses, with new data continuously complementing earlier datasets.

Ideally, molecular species-delimitation should be based on: (i) nuclear and mitochondrial markers, to reflect gene flow of both nuclear and mitochondrial genes; (ii) genomic scale for nuclear markers to cover numerous independent loci; (iii) sufficiently variable to capture recently diverged species; (iv) repeatable, so that datasets can be complemented once more material is available; (v) universal to the extent that datasets can complement each other. In 2012, ultraconserved elements (UCEs) were introduced as a quick and essentially universal way to obtain "thousands of genetic markers spanning multiple evolutionary timescales" (Faircloth et al., 2012). UCEs appear to fulfil many of the above mentioned requirements for nuclear markers. However, whether they harbor enough variation to capture divergence among recently diverged species remains an open question, since by definition they are highly conserved.

In this study, we examine the use of UCEs for species delimitation in Central European bees. We include examples of both putative mitochondrial introgression and of multiple "barcodes" per species, investigating how UCEs can overcome the main drawbacks for species delimitation using DNA barcoding developed above. We focused on the following European species complexes: *Andrena amieti/allosa/bicolor/montana* ; *Andrena barbareae/cineraria; A. dorsata/propinqua; A. carantonica/trimmerana/rosae; Lasioglossum al-*

pigenum/bavaricum/cupromicans; Nomada goodeniana/succincta . Mitochondrial introgressions have been suggested for four of these cases (Schmidt et al., 2015; see details below); low-divergence were suggested for the controversial *A. carantonica /trimmerana*complex (Schmidt et al. 2015); while deep within-species divergences not associated with morphological differentiation have been documented the*Andrena amieti/allosa/bicolor/montana* clade (Praz et al. 2019). Most of these cases are also controversial with respect to morphological delimitations, so that current evidence based on the combined characteristics of morphology and COI-based DNA barcodes does not enable definite conclusions on the status of these species.

2. Material and Methods

2.1. Species complexes

Six species complexes showing discrepancies between morphological and COI-based identifications in the Swiss bee fauna were selected, mainly based on the comprehensive study of Schmidt et al. (2015). All reported cases of mitochondrial introgression in Switzerland were included, with the exception of *Andrena nitida/limita* and *Colletes hederae/succinctus* , for which not enough material was available. The following section provides information on the six species complexes.

2.1.1. Species complex 1: Andrena allosa/amieti/bicolor/montana

Species delimitation within this group (hereafter "*bicolor* -group") has long remained controversial, especially in the alpine region where two taxa with debated status (F Amiet, Hermann, Müller, & Neumeyer, 2010) co-occur with *A. bicolor* , namely*A. allosa* and *A. montana* . Phylogenetic analyses on a mitochondrial (COI) and a nuclear gene (LW-rhodopsin) confirmed the taxon validity of both *A. allosa* and *A. montana* , but also revealed the presence of a new, until then undescribed cryptic species,*A. amieti* (Praz et al., 2019). These analyses also revealed two sympatric clades for *A. bicolor* (also reported in Schmidt et al., 2015) and two sympatric clades for *A. amieti* . Genetic distances between these sympatric clades were comparatively high, with approximately 3.7% and 2.4% for *A. bicolor* and *A. amieti* , respectively, and hence comparable to distances among valid species in this group. In addition, the two sympatric clades within*A. amieti* formed a paraphyletic unit from which the distinct species *A. allosa* arose. Whether the mitochondrial clades found in *A. bicolor* or *A. amieti* represent additional cryptic species within this group, remains unclear.

To complement the dataset, two specimens collected in Greece and probably representing an undescribed species were included in this species complex; this species is referred to as *Andrena* sp3, as in Praz et al. (2019).

2.1.2. Species complex 2: Andrena barbareae/cineraria

A. barbareae and *A. cineraria* are sibling species, morphologically very close, although identifiable in most cases by a combination of characters in both genders (Amiet et al. 2010). *A. cineraria* has a wider distribution than *A. barbareae* , which is mainly restricted to the Alps. Both species are polylectic but exhibit different phenologies, with two generations for *A. barbareae* and one for *A. cineraria* . Because of their morphological, biogeographical and phenological differences, both species are generally considered as separated although this view was recently challenged because both taxa share identical barcodes (Schmidt et al., 2015).

2.1.3. Species complex 3: Andrena carantonica/trimmerana/rosae

Taxonomical delimitation in this species complex is a long-standing enigma with controversial species delimitation hypotheses due to morphologically divergent generations and unclear morphological differentiation. First, *Andrena rosae* and *A. stragulata*are considered by most authors to represent the summer and spring generations of the same species (Falk, 2016; Reemer, Groenenberg, Van Achterberg, & Peeters, 2008; Westrich, 2014). Nevertheless, both taxa exhibit differences in terms of morphology, pollen collecting behaviour and,

possibly, nesting sites (Amiet et al., 2010; van der Meer, Reemer, Peeters, & Neve, 2006; Westrich, 2014). Second, morphological differentiation of *A. carantonica* and *A. trimmerana* is challenging. No clear morphological character allows to separate females; and while males of the first generation of *A. trimmerana*differ in the morphology of the mandible, those of the summer generation cannot be differentiated from those of *A. carantonica* . Both taxa exhibit distinct phenologies with only one generation for *A. carantonica* (April to May) and two for *A. trimmerana*(March-April, June-July), although isolated late-summer specimens of*A. carantonica* are known (C. Praz and R. Paxton, unpublished data). Both taxa overlap in their distribution area, but *A. carantonica* is much more abundant than *A. trimmerana,* for which only a few occurrences were reported in Switzerland or Germany.

2.1.4. Species complex 4: Andrena dorsata/propinqua

Depending on the author, *A. dorsata* and *A. propinqua* are considered separate or conspecific taxa (Amiet et al., 2010; Gusenleitner & Schwarz, 2002; Schmid-Egger & Scheuchl, 1997). Morphologically, the separation of both taxa is complicated and subject to errors, especially at the European scale. At the genetic level, two distinct clades mostly corresponding to the two taxa were recovered, however, both species were previously found to share COI, suggesting that identification errors could underlie the observed barcode sharing (Schmidt et al., 2015).

2.1.5. Species complex 5: Lasioglossum alpigenum/bavaricum/cupromicans

Species delimitation in this complex is generally accepted based on clear differences in male genital morphology (Amiet, Herrmann, Müller, & Neumeyer, 2001; Edmer, 1970). Identification of females is however challenging, and *L. bavaricum* and *L. cupromicans* were recently suggested to share the same COI barcode. However, since no male of *L. bavaricum* had been sequenced, identification of the sequenced specimens for this taxon remain tentative.

2.1.6. Species complex 6: Nomada goodeniana/succincta

The morphological separation between these two species is mainly relying on colour patterns and is therefore prone to identification errors, although both appear to differ in their hosts, phenology and possibly in the chemical composition of mandibular glandular secretions (Kuhlmann, 1997). Schmidt et al. (2015) found two divergent clusters for *N. succincta* : a northern European cluster containing specimens of *N. goodeniana* and a southern European cluster. A similar result was found in England (Creedy et al., 2019). As described above for *A. dorsata/propinqua* , the reported COI barcode sharing between *N. goodeniana* and *N. succincta* could potentially be due to misidentification, at least in Germany, where a previous study suggested that both species did not share COI barcode (Diestelhorst & Lunau, 2008).

2.2. Sampling

Most specimens were sampled across Switzerland in the frame of the "Red List of Swiss bee" project between 2008 and 2019. Additional specimens were collected in France, Italy and Greece. For the *A. bicolor*complex, sites known to harbour large populations or several species/clades in sympatry were additionally sampled in 2018. Bees collected within the red list surveys were killed in ethyl acetate, pinned and preserved dry. Samples collected in 2018 were preserved in 70% EtOH at 4°C to ensure good DNA preservation. All bees were morphologically identified by one of us (C. Praz); for *A. carantonica/trimmerana* , phenology was used in addition to morphology for identification. Further information on the sampling, as well as metadata and sampling maps are provided as Supplementary information (S1-S2); in addition, the COI sequence of every specimen used in this study has been uploaded onto the BOLD platform.

2.3. COI sequencing and phylogeny

The cytochrome oxidase unit I (COI) barcode fragment of all specimens was sequenced either by Sanger or NGS-barcode sequencing using the primer pairs LepF/LepR (Hebert, Penton, Burns, Janzen, & Hallwachs, 2004) or mlCOIntF/HCO (Leray et al., 2013) (following protocols in Gueuning et al., 2019). Raw sequences were imported into Geneious v11.0.5 and consensus sequences between forward and reverse sequences were constructed for each specimen using the Geneious assembler. Consensus sequences were aligned per species complex using MAFFT v7.308 (Katoh & Standley, 2013). The absence of stop codons within sequences was confirmed by translating and skimming the consensus sequences. Phylogenetic trees were built with RAxML v8.2.11 (Stamatakis, 2014) using the GTR GAMMA model and 100 bootstrap replicates. Since sequences were obtained using different methods (i.e. Sanger, NGS) and different primers, sequences were truncated to the same length (i.e. 265 bp).

2.4. UCE library preparation

Whole body DNA extractions were performed overnight in a proteinase K buffer at 56°C and purified using a Qiagen Biosprint 96 extraction robot following the manufacturer's protocol. Extracts were quantified using Qubit v4 (Thermofisher Scientific) and 50 ng DNA per specimen were sonicated to 500 bp fragment length using a Bioruptor ultrasonicator (Diagenode). Two independent dual-indexed libraries containing each 96 specimens were constructed using a Kapa Hyper prep kit (Roche) using one fourth of the manufacturer's recommended volumes (as described in Branstetter, Longino, Ward, & Faircloth, 2017). PCR amplifications were performed in the recommended volumes. PCR products were quantified using a Qubit v2 and each row of a 96-well PCR plate were pooled equimolarly (i.e. for total of 8 pools). Libraries were UCE enriched using the Hymenopteran v2 hybridization kit (UCE Hymenoptera 2.5Kv2 Principal/Full, myBaits, Arborbiosci). Each enrichment was performed on a single pool of 12 individuals using 500 ng. The enrichment protocol followed the manufacturer's recommendations with a hybridization step of 24 h at 65°C, followed by a PCR amplification with 14 cycles. Pools were sequenced on a Miseq using five Illumina v3 kits (2 x 300 bp; Illumina, location, Switzerland).

2.3. Bioinformatic processing of UCE data

Fastq reads were demultiplexed on the Miseq and data from all runs were merged and processed mainly using PHYLUCE tools (Faircloth, 2016). Raw data were cleaned with illumprocessor (Faircloth, 2016), a tool wrapped around trimmomatic (Bolger, Lohse, & Usadel, 2014). Clean reads were assembled with SPAdes v3.12.0 (Nurk et al., 2013) using the single-cell flag ("–sc"), carful option ("–careful") and a coverage cutoff value of five ("–cov-cutoff"). Obtained contigs were mapped against the corresponding UCE reference file using Lastz (Harris, 2007) and matching reads were extracted and aligned by species complex using MAFFT (Katoh & Standley, 2013). Since species within each complex are relatively close related ($<$ 30-50 MYA), alignments were edge-trimmed and not internal-trimmed (Faircloth, 2016). Loci shared by less than 75% of the maximum number of specimens sharing a locus were filtered out. Remaining alignments were concatenated and saved in fasta format. An additional filtering step was applied to remove specimens with more than 90% missing data.

2.4. UCE analyses

The remaining concatenated alignments of the UCE amplicons were used for phylogenetic analyses. Maximum likelihood trees were produced for each species complex with RAxML v8.2.11 using the same parameters as for the COI RAxML trees. For comprehensiveness, two separate trees were produced for the "*bicolor* -goup" (i.e. one for *A. amieti* and one for *A. bicolor*). Genetic distances between species/lineages for each species complex were computed with MEGA-x (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) using the Tajima-Nei model.

Multivariate analyses and genetic distance tests were conducted in R, mainly using the adegenet package (Jombart, 2008). Sequences were first imported into R using the "*fasta2genlight* " function which reads aligned sequences and extracts binary SNPs before converting files into a genlight object. After conversion, the SNPs matrices were filtered to retain only variable sites containing less than 15% missing data. Datasets were then screened for significant departure from Hardy-Weinberg equilibrium using the dartR package (Gruber, Unmack, Berry, & Georges, 2018). Principal component analyses were performed using the "*dudi.pca* " function (ade4 package; Dray & Dufour, 2007) without scaling or centering the data. PCA results were plotted with ggplot2 (Wilkinson, 2011) using the two first components. To further identify and describe genetic clusters, discriminant analyses of principal components (DAPC) were performed. A first approach was used to verify the group's membership using a prior knowledge on the species assignments. For the "*bicolor* -group", *A. bicolor* and*A. amieti* were divided into two distinct groups (e.g. mitochondrial lineage 1 and 2; hereafter referred to as ML1 and ML2). The optimal number of PCs to retain was identified using both the plotted cumulative variance of the eigenvalues and a cross-validation method implemented in the "*optim.a.score* " function. Results of posterior membership probabilities for each specimen were plotted using ggplot2. In a second approach, we ran a DAPC by grouping specimens into genetic clusters without species a priori knowledge. The function "*find.clusters* " was used to determine the optimal number of genetic clusters which was defined as the solution harbouring the lowest Bayesian Information Criterion (BIC) value. Further, we computed for each species complex pairwise fixation indexes (Fst) between putative species (with *A. amieti* and *A. bicolor* treated as two lineages) using the dartR package with 10,000 permutations. Levels of observed genetic differences were tested using analyses of molecular variance (AMOVAs), using the dartR package with 10,000 permutations.

Because a potential pattern of isolation by distance (IBD) was observed in the phytogenic tree of *A. amieti* , correlation between genetic and geographical distance between sampling locations was computed using the "gl.ibd" function (dartR package) with 10,000 permutations. Genetic distances were computed upon Nei's genetic distance (Nei, 1972) using the "*stamppNeisD* " function from the StAMPP package (Pembleton, Cogan, & Forster, 2013) and geographical distances according to the "haversine method" using the geosphere package (Hijmans, Williams, & Vennes, 2019).

Finally, three independent analyses were performed for testing species delimitation. (I) First, we performed a Generalized Mixed Yule Coalescent model (GMYC) on ultrametric trees. Trees for each species complex were built with BEAST2 v2.5.2 (Bouckaert et al., 2014) using the JC69 substitution model and a strict molecular clock with a fixed rate of 1.0. Priors followed a yule model with a uniform distribution for "birthRate". MCMC ran for 250 million generations with sampling every 1000 generations. Chain convergence was assessed using the software TRACER v1.6 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018). For computational purposes, trees were resampled to a total of 100 trees using the logCombiner software. The ultrametric trees were then imported into R using the *ape* package (Paradis & Schliep, 2019). GMYC was performed on the last tree using the *splits* package (Ezard, Fujisawa, & Barraclough, 2009). Interval of species number was set between 0 and 10 and the analysis was run using the single-threshold version. (II) Second, results from the first analyses were cross-validated using a Bayesian implementation of the GMYC model (bGMYC) (Reid & Carstens, 2012). For each species group, the maximum number of possible "species" was set as twice the number of expected species present in the species complex. The MCMC was set to 11,000 generations (1000 generations burnin) and sampled every 10 generations ("thinning"). (III) Third, we performed an analysis on the concatenated sequences using the Bayesian Phylogenetics and phylogeography model (BPP) (Yang & Rannala, 2010). The BPP analyses were ran using the A11 model (i.e. unguided species delimitation analysis; Yang & Rannala, 2014) on the nexus files (each corresponding to one UCE) obtained after the 75% threshold filtering step. The population file was designed so that specimens were assigned to their species. For the "*bicolor* -group", based on the phylogeny trees,*A. bicolor* was divided into two distinct lineages and *A. amieti* in one. Alpha and beta parameters of the inverted gamma distribution of the theta prior (average proportion of different sites between two sequences) were set to 3 and 0.004, respectively. For the tau prior, alpha and beta were respectively set to 3 and 0.002. The analyses was run two times with a MCMC of 50,000 generations and with a 10% burnin period.

3. Results

3.1. UCE sequencing

All combined, the Miseq runs produced on average 526,308 (SD ± 356,015) reads per specimen. The average number of loci per specimen after assembly was 7398 (SD ± 5099). After retaining only loci matching the UCE reference file and filtering for a 75% matrix completeness, the number of retrieved loci varied between 686 and 1860 across the five species complexes (Supplementary table S3). Resulting concatenated reads were on average 899,690 bp long; the shortest reads were obtained for *N. goodeniana/succincta* (i.e. 429,820 bp) and the longest for *A. amieti/bicolor* (i.e. 1,459,230 bp). In total, seven specimens did not pass the 90% missing data filter (Supplementary information S1) and were therefore excluded from downstream analyses. No significant departure form Hardy-Weinberg equilibrium was observed. The bi-allelic SNPs screening recovered on average 18,545 bi-allelic SNPs (SD ± 23,430) across all species complexes (Supplementary information S3).

3.2. Comparison of mitochondrial and nuclear analyses

3.2.1 Species complex 1: Andrena allosa/amieti/bicolor/montana

The phylogenetic trees performed on the COI and UCEs datasets provided very similar topologies for *A. bicolor* (Figure 1). Both trees showed distinct monophyletic clades with strong bootstrapping support (hereafter BS values) for the UCE tree. Only one specimen (i.e. "GBIFCH00135933") sampled in Southern France was misplaced in the UCE tree compared to its position in the corresponding mitochondrial tree. For *A. amieti* , all specimens formed one monophyletic clade in the UCE tree, contrasting with the two paraphyletic clades in the COI tree. Furthermore, there was no apparent structuring in the UCE tree among mitochondrial lineages sampled in the alpine region. Specimens with large amounts of missing data ([?] 60%) exhibited longer branches in the UCE tree (e.g. "GBIFCH00131686", "GBIFCH00136250"). Strong isolation by distance was found between the southern Italian and Alpine populations (R^2 = 0.7221, p-value = 0.039; Supplemental information S4). These two individuals formed a distinct monophyletic clade sister to all alpine specimens in the UCE phylogenetic tree.

The DAPC with a prior knowledge on species assignments correctly reassigned membership for the majority of specimens (Supplementary information S5). The plotted cumulative variance of the eigenvalues suggested to retain the first eight principal components (conserving 70.3% of the total variance; Supplementary information S6). All specimens were correctly reassigned with a 100% membership probability for three taxa (i.e. *A. allosa* , *Andrena* sp3, *A. montana*). For *A. bicolor* ML1, only one specimen (i.e. "GBIFCH00135933" from Southern France) revealed mixed membership probability, with 45.64% posterior probability (hereafter "pp") to belong to ML1 and 54.36% to ML2. For *A. bicolor* ML2, one specimen (GBIFCH00117401) also showed mixed membership with a slight probability (i.e. 3.8%; Supplementary information S5) of belonging to *A. bicolor* ML1. The mixed membership probability for those two specimens are congruent with the PCA results where both specimens are found marginally away from the main *A. bicolor* ML2 aggregation (Supplementary information S7).

Between both lineages of *A. amieti* , genetic cluster assignments were much less supported and only the two specimens sampled in south Italy were assigned with a 100% probability to ML1. The lack of a clear separation between mitochondrial lineages for the alpine specimens is suggesting considerable levels of admixture.

The AMOVA (Table 1) depicted strong genetic difference (i.e. 43.43%; p-value [?] 0.0001) between the two *A. bicolor* lineages but no significant difference between the two *A. amieti* lineages. The lowest, yet significant fixation index (Table 2) was obtained between both lineages of *A. bicolor* (Fst = 0.138). Nei's genetic distance between both lineages within *A. bicolor* (i.e. 0.00231) was slightly higher that between *A. allosa* and *A. amieti* ML1 and ML2 (i.e. 0.00261 and 0.00299, respectively).

The GMYC analysis computed on all specimens identified nine clusters (Supplementary information S8): two clusters corresponding to the mitochondrial lineages found within *A. bicolor* , one cluster with with *A. amieti* and *A. allosa* merged together, and six clusters composed of only one specimen. The bGMYC analyses identified the same 9 clusters with a high probability (p = 0.95 - 1; Supplementary information S8). All other scenarios had very low posterior probabilities (pp = 0 - 0.05). The two parallel *BPP* analyses converged and were highly congruent (Supplementary information S9). Both runs depicted: (i) one tree model [((*A. allosa, A. amieti*ML1+ML2), ((*A. bicolor* ML1, *A. bicolor* ML2), *Andrena sp3*))] with a posterior probability of [?] 0.99; (ii) 5 delimited species (i.e. *Andrena sp3* , *A. bicolor* ML1, *A. bicolor* ML2, *A. allosa* , *A. amieti* ML1+ML2), all with a posterior probability of 1; (iii) and a posterior probability of 1 for having 5 species present in the dataset. Finally, the DAPC analyses performed without a prior knowledge on species identifications identified K = 3 and K = 4 as best solution for *A. amieti* ML1+ML2 and *A. bicolor* ML1+ML2, respectively. Table 3 summarized the number of clusters found for each analysis.

3.2.1. Species complex 2: Andrena barbareae/cineraria

Mitochondrial and nuclear phylogenies were discordant, with no clear separation between both species in the COI tree and two well-supported monophyletic clades corresponding to the two morphological species in the UCE tree (100% BS values; Figure 1). Results from the UCE phylogenetic tree were corroborated by the PCA in which both species were clearly separated (Supplementary information S7). The DAPC analyses correctly reassigned membership for all specimens with 100% probability (Supplementary information S5). The AMOVA revealed that 52.74% of the total observed variance could be explained by the species level. The GMYC and bGMYC provided similar results (Supplementary information S8). Both analyses grouped all specimens morphologically identified as*A. cineraria* in one clade. For *A. barbareae* , both analyses suggested the presence of three distinct clades. Both BPP runs were congruent and highly supported the presence of three species (pp = 1; Supplementary information S9). The runs however disagreed with respect to the phylogenetic relationships among the three species. The first run depicted three different possibilities for the species trees, with the most likely (pp = 1) tree being: [(*A. barbareae* , (*A. cineraria* , *A. vaga))*]. The second run depicted only one tree [((*A. barbareae* , *A. cineraria)* , *A. vaga)],*corresponding to the expected tree based on the phylogeny. In the first run this solution was supported at 75%. Congruent with the BPP analyses, the DAPC identified K = 3 (with outgroup) as the best solution (Supplementary information S10, Table 3). Clustering of all specimens corresponded to the morphological identifications.

3.2.2. Species complex 3: Andrena carantonica/trimmerana/rosae

Mitochondrial phylogenies recovered well-supported (BS 72-100%) monophyletic clades for *A. carantonica* and *A. rosae* , but not for *A. trimmerana* , which was composed of two clades forming a paraphyletic unit from which *A. carantonica* arose. One clade was composed of two specimens of *A. trimmerana* sampled in western Switzerland and was sister to the *A. carantonica* clade; support for this sister relationship was high (BS 93%; Figure 1). In contrast, all three species appeared as strongly supported monophyletic clades in the UCE tree ([?] 90% BS; Figure 1). Spring and summer generations of*A. rosae* and of *A. trimmerana* were intermixed in both mitochondrial and UCE trees, supporting the view that *A. stragulata* and *A. spinigera* constitute the morphologically differentiated spring generation of *A. rosae* and *A. trimmerana* , respectively. Genetic distance between *A. carantonica* and *A. trimmerana* was relatively low (Nei's D = 0.00061), although AMOVA and pairwise Fst depicted significant difference between both species (Table 1-2). The PCA with all three species showed no difference between *A. carantonica* and *A. trimmerana* , however when removing *A. rosae* from the analyses, both species were separated on the first two components (Supplementary information S7). The GMYC and bGMYC analyses failed to separate both*A. carantonica* and *A. trimmerana.* In the DAPC both species were also not separated with K = 2 and K = 3 but were with K = 4 (Supplementary information S10). All three clustering scenarios had very closed BIC values. The BPP analyses however highly supported the presence of three distinct species, with the following tree topology [((*A. carantonica* , *A. trimmerana*), *A. rosae*)].

3.2.3. Species complex 4: Andrena dorsata/propinqua

Strong mito-nuclear discordances were observed within this species complex. In mitochondrial trees, Swiss specimens formed two clusters corresponding to morphological identifications (Figure 1), but one specimen of *A. propinqua* (GBIFCH00133244) collected in southern France was sister to a well-supported clade containing all other specimens of *A. dorsata* and of *A. propinqua* (Figure 1). Our sampling also included one specimen of *A. dorsata* from this French site (GBIFCH00133243). Phylogenetic trees and PCAs based on UCEs recovered both species as separated clusters (Figure 1, Supplementary information S7); the French specimens were not particularly divergent. Both GMYC and bGMYC analyses, the BPP analyses and DAPC analysis successfully separated both species (Table 3, Supplementary information S8-S10). Taken together, these results indicate that *A. dorsata*and *A. propinqua* are valid species.

3.2.4. Species complex 5: Lasioglossum alpigenum/bavaricum/cupromicans

Comparison of mitochondrial and nuclear trees revealed mitochondrial-nuclear discordance for *L. bavaricum* and *L. cupromicans* (Figure 1). Both taxa were well delimited with highly supported monophyletic clades (95% bootstrap value) in the UCE tree but shared the same COI barcode. All *L. alpigenum* specimens clustered in a single monophyletic clade sister to both other taxa in both trees. Beside the GMYC analysis that over-clustered *L. bavaricum* and*L. cupromicans* (Supplementary information S8), all other analyses were congruent (Table 3) and supported the hypothesis of three distinct species as previously postulated based on morphology.

3.2.5. Species complex 6: Nomada goodeniana/succincta

COI and UCE trees depicted two well defined monophyletic clades (Figure 1). The bootstrap support for monophyly of *N. goodeniana* was low in the mitochondrial trees due to the presence of two slightly divergent specimens of *N. goodeniana* collected south of the Alps. In the nuclear trees, these two specimens clustered with other specimens of*N. goodeniana* with high support values. The species delimitation tests also highly supported the hypothesis of two separated species (Table 3, Supplementary information S8-S10).

4. Discussion

4.1. Ultraconserved elements successfully delimit species in all investigated cases

In all six complexes of wild bee species examined here, UCEs provided robust species hypotheses and clearly outperformed COI for species delimitation. The main results of our study can be summarized as follows: First, we provide strong evidence of mitochondrial introgression in two species pairs (*Andrena barbareae* and*A. cineraria* , *Lasioglossum cupromicans* and *L. bavaricum*): UCEs were in agreement with morphology but not with COI, which suggests that barcode sharing occurs in these species pairs. Second, three species complexes presented multiple mitochondrial DNA barcodes in a single biological species (i.e. *Andrena amieti* ,*A. propinqua* , *A. trimmerana*); for all three species UCEs recovered strongly supported monophyletic groups which were in agreement with morphology, while the two mitochondrial barcodes within each species formed a paraphyletic assemblage from which another species arose (*A. allosa* , *A. dorsata* and *A. carantonica* , respectively), resulting in the absence of a barcode gap and unresolved mitochondrial species delimitation. Third, our results suggest that the two mitochondrial clades observed within *Andrena bicolor* probably represent two distinct cryptic species. In addition, UCE-based species delimitation solved long-standing controversies in the taxonomy of Central European bees; in particular, the two generations within each of*A. rosae* and *A. trimmerana* , respectively, which do not appear to represent distinct species; and the distinctiveness of the other species pairs or triplets investigated here, which is strongly confirmed by the UCE data.

4.2. DNA barcoding errors

COI-based barcoding is subject to two types of errors. The first error (similar to type I error) occurs when one biological species is associated with two distinct DNA barcodes, as observed for *A. amieti* , *A. propinqua* and *A. trimmerana* . Type I errors ultimately lead to erroneous detection of two hypothetical species within a single biological species. Most often, these errors are triggered by deep within-species divergences or artefacts such as nuclear insertions (Song, Buhay, Whiting, & Crandall, 2008). The second error (i.e. type II error) occurs when DNA barcoding fails to recognize two distinct species because of barcode sharing, as observed between the pairs *Andrena barbareae/cineraria* and *Lasioglossum cupromicans /barvaricum* .

Identifying the exact biological mechanism behind these barcoding errors can be tedious, but often they are linked to incomplete lineage sorting, hybridization followed by introgressions, demographic disparities, *Wolbachia* infections or sex-biased asymmetries (i.e. male-biased dispersal, mating behaviour or sex-biased offspring production) (Toews & Brelsford, 2012). Most often these events occur in recently diverged species and are not necessarily mutually exclusive (Mutanen et al., 2016). In this study, the low number of specimens sampled and sequenced render the investigation on the underlying mechanism difficult. A more complete sampling across the entire distribution would be necessary to separate incomplete lineage sorting from the other mechanisms. Indeed, incomplete lineage sorting is most often not associated with any biogeographical pattern (Funk & Omland, 2003; Toews & Brelsford, 2012). In contrast, events such as hybridization/introgression often leave biogeographical footprints because they are unidirectional, which implies that the gene flow is directed from the native taxon towards the colonized taxon (Currat, Ruedi, Petit, & Excoffier, 2008; Nevado, Fazalova, Backeljau, Hanssens, & Verheyen, 2011; Pons et al., 2014). Therefore, introgression levels are highest at the hybridization zone and fade away over the colonized distribution zone (Toews & Brelsford, 2012). Further work with a wider geographic coverage would be necessary to unravel the cases of DNA barcoding errors documented here.

4.3. Could the UCEs have overlooked additional levels of cryptic diversity?

With regard to the low rate of evolution of UCEs, an important question in our study and more generally with the use of UCEs for species delimitation is whether they can successfully uncover variation between recently diverged species. It could be argued that the cases of mitochondrial paraphyly (i.e. *A. amieti* , *A. trimmerana* and *A. propinqua*) in fact represent additional, overlooked instances of cryptic species, and that the UCEs rate of evolution is too low to recover these divergences. At least for *A. amieti* , our sampling across the entire known distribution of this species enables us to exclude this scenario. We included specimens from the Alps and from the Apennines in Southern Italy, some 600 km from the nearest Alpine population; the Apennine specimens are morphologically slightly divergent from the Alpine populations (Praz et al. 2019). In the COI tree (Figure 1), the southern Italian specimens all clustered in one mitochondrial clade, while the Alpine specimens were distributed over both mitochondrial clades. The UCEs recovered two strongly supported clades within *A. amieti* , one corresponding to the Southern Italian population and the other including all alpine specimens (Figure 1). This result strongly contradicts the hypothesis of two separated lineages corresponding to both mitochondrial clades. Rather, UCE results agree with the strong geographic separation of the Alpine and Apennine populations and with their slight morphological differentiation.

In the two other cases of mitochondrial paraphyly (i.e. *A. trimmerana* and *A. propinqua*) investigated in our study, the presence of additional cryptic species can not completely be rejected. We however deem this scenario as strongly unlikely since near-cryptic species in bees are almost exclusively associated with some level of morphological differentiation in highly variable character such as pile colour or punctuation (McKendrick et al., 2017; Pauly et al. 2019; Praz et al. 2019). In our study, such morphological variations were not observed in the divergent specimens (it was admittedly also not observed between the two clades within *A. bicolor* , although more specimens of both genders and both generations are needed to address this question thoroughly). In addition, these divergent specimens in mitochondrial trees where nested within the clades of conspecific specimens in the UCE trees. We speculate that such high within-species divergences in

mitochondrial barcodes will prove more common than previously expected once barcoding with continental-scales sampling will be achieved (Hinojosa et al., 2019; Schmidt et al., 2015).

4.4. Comparison of different species delineation methods

The method of species delimitation that provided the results most in agreement with current morphological hypotheses was BPP. By contrast, results by both (b)GMYC were less congruent with morphological species hypotheses, and in several cases had the tendency to inflate the number of species. Compared to BPP, GMYC analyses can overestimation or underestimations species delimitation rates (Carstens et al., 2013; Luo, Ling, Ho, & Zhu, 2018), especially in the presence of high intraspecific variation (Talavera, Dincă, & Vila, 2013). In our particular case, specimens with low-quality input DNA yield high levels of missing values, which led to longer branches in the trees. In most cases, the GMYC analyses split specimens harboring long branches singleton species (Supplementary information S8) which ultimately inflated the overall species number. Therefore, GMYC analyses should be interpreted with caution when applied on UCE data.

4.5. Concluding remarks on the use of UCEs for species delimitation

Our results confirm that UCEs can provide sufficient variation at shallow time scale in insects to enable species discrimination, adding to previous evidence gathered in vertebrates (Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016; Zarza et al., 2018). Harvey et al. (2016) comprehensively compared the utility of sequence capture methods, specifically using UCEs as in our study, and RAD-Seq for shallow phylogenies. They found that both techniques resulted in similar phylogenetic hypotheses and branch support values; and that RAD-seq provided more overall information while sequence-capture provided higher per-locus-information. They also suggested that the high amount of information typical of RAD-seq was not necessarily an advantage when the inherent question was phylogeography, phylogeny or species delimitation. Harvey et al. (2016) concluded that sequence capture is more useful in systematics because of its repeatability, the possibility to use low-quality samples, the ease in read orthology assessment, and the higher per-locus information.

Our results build upon this early work and largely confirm these predictions. RAD-seq datasets would have been nearly impossible to gather for the species investigated here due to low DNA quality or quantity. The possibility of processing specimens belonging to three different families simultaneously, and to iteratively assemble datasets, represent particularly promising advantages of UCE capture methods for species delimitation. Future work should focus on very recently diverged taxa to further determine the level of divergences that can be recovered with these conserved markers. In addition, whether UCEs will enable the detection of hybrids, and to what extend the presence of these hybrids impact the tree reconstruction or the species delimitation analyses should be investigated. Lastly, our analyses strongly suggest the presence of two cryptic species within one of the most common European bee, *Andrena bicolor* . Enlarging our dataset to the entire geographical range of *A. bicolor* will be necessary to further untangle this remarkable case of cryptic species in bees.

Acknowledgements

We are grateful to all collaborators who provided specimens for this study. We would also like to thank Silas Bossert, Gideon Pisanty, Robert Paxton, Daniel Frei and Beatrice Frey for there valuable repective contribution in regards to bioinformatic, laboratory and statictical aspects of this study. This work was mainly funded by the Swiss Federal Office for Agriculture (FOAG); the sampling of bees was funded by the Swiss Federal Office for Environment (FOEN).

References

Amiet, F., Hermann, M., Müller, A., & Neumeyer, R. (2010). *Apidae 6, Andrena, Melliturga, Panurginus, Panurgus* (Fauna Helv). CSCF & SEG.

Amiet, F., Herrmann, M., Müller, A., & Neumeyer, R. (2001).*Apidae 3: Halictus, Lasioglossum. Fauna Helvetica 6, 208 pp. In German and French.* (Fauna Helv). CSCF & SEG.

Andriollo, T., Naciri, Y., & Ruedi, M. (2015). Two mitochondrial barcodes for one biological species: The case of European Kuhl's pipistrelles (chiroptera). *PLoS ONE* . doi: 10.1371/journal.pone.0134881

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

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* ,*10* (4), e1003537. doi: 10.1371/journal.pcbi.1003537

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. doi: 10.1111/2041-210X.12742

Brunner, P. C., Fleming, C., & Frey, J. E. (2002). A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach. *Agricultural and Forest Entomology* , *4* (2), 127–136. doi: 10.1046/j.1461-9563.2002.00132.x

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology* . doi: 10.1111/mec.12413

Creedy, T. J., Norman, H., Tang, C. Q., Qing Chin, K., Andujar, C., Arribas, P., ... Vogler, A. P. (2019). A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding . In *Molecular Ecology Resources* . doi: 10.1111/1755-0998.13056

Dellicour, S., & Flot, J. F. (2018). The hitchhiker's guide to single-locus species delimitation. *Molecular Ecology Resources* . doi: 10.1111/1755-0998.12908

Diestelhorst, O., & Lunau, K. (2008). Contribution to the clarification of the species status of Nomada goodeniana. *Entomologie Heute* , 165–171.

Dray, S., & Dufour, A. B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* . doi: 10.18637/jss.v022.i04

Dupuis, J. R., Roe, A. D., & Sperling, F. A. H. (2012). Multi-locus species delimitation in closely related animals and fungi: One marker is not enough. *Molecular Ecology* , *21* (18), 4422–4436. doi: 10.1111/j.1365-294X.2012.05642.x

Edmer, A. W. (1970). Die Bienen des Genus Halictus LATR. S.L. im Großraum von Linz (Hymenoptera, Apidae) Teil II. *Naturk.Jb.Stadt Linz* , 19–82.

Ezard, T., Fujisawa, T., & Barraclough, T. G. (2009). Splits: species' limits by threshold statistics. *R Package Version* .

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

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. doi: 10.1093/sysbio/sys004

Falk, S. (2016). Field Guide to the Bees of Great Britain and Ireland.*Bee World* , *93* , 85. doi: 10.1080/0005772X.2016.1257474

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* . doi: 10.1111/1755-0998.12745

Gueuning, M., Ganser, D., Blaser, S., Albrecht, M., Knop, E., Praz, C., & Frey, J. E. (2019). Evaluating next-generation sequencing (NGS) methods for routine monitoring of wild bees: Metabarcoding, mitogenomics or NGS barcoding. *Molecular Ecology Resources* , *19* (4), 847–862. doi: 10.1111/1755-0998.13013

Gusenleitner, F., & Schwarz, M. (2002). Weltweite Checkliste der Bienengattung Andrena mit Bemerkungen und Ergänzungen zu paläarktischen Arten (Hymenoptera, Apidae, Andreninae, Andrena). *Entomofauna Supplement* , *12* , 1–1280.

Harris, R. S. (2007). Improved Pairwise Alignment of Genomic DNA. PhD thesis, Pennsylvania State University. doi: 10.1016/j.brainres.2008.03.070

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology* . doi: 10.1093/sysbio/syw036

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* , *270* (1512), 313–321. doi: 10.1098/rspb.2002.2218

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator.*Proceedings of the National Academy of Sciences* , *101* (41), 14812–14817. doi: 10.1073/pnas.0406166101

Hijmans, R. J., Williams, E., & Vennes, C. (2019). geosphere: Spherical Trigonometry. R package version 1.5-10. *Package Geosphere* .

Hinojosa, J. C., Koubínová, D., Szenteczki, M. A., Pitteloud, C., Dincă, V., Alvarez, N., & Vila, R. (2019). A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly Thymelicus sylvestris. *Molecular Ecology* , mec.15153. doi: 10.1111/mec.15153

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* . doi: 10.1093/bioinformatics/btn129

Karanovic, T., Djurakic, M., & Eberhard, S. M. (2016). Cryptic Species or Inadequate Taxonomy? Implementation of 2D Geometric Morphometrics Based on Integumental Organs as Landmarks for Delimitation and Description of Copepod Taxa. *Systematic Biology* . doi: 10.1093/sysbio/syv088

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

Klopfstein, S., Kropf, C., & Baur, H. (2016). Wolbachia endosymbionts distort DNA barcoding in the parasitoid wasp genus Diplazon (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society* , *177* (3), 541–557. doi: 10.1111/zoj.12380

Kuhlmann, M. (1997). Zum taxonomischen Status von Nomada goodeniana (KlRBY, 1802) und Nomada succincta PANZER, 1798 (Hymenoptera, Apidae).*Entomofauna* , *18* (32), 521–528.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms.*Molecular Biology and Evolution* . doi: 10.1093/molbev/msy096

Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: rules and exceptions. *Journal of Biological Research-Thessaloniki* , *24* (1), 2. doi: 10.1186/s40709-017-0060-4

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., ... Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution* . doi: 10.1002/ece3.4905

Leneveu, J., Chichvarkhin, A., & Wahlberg, N. (2009). Varying rates of diversification in the genus Melitaea (Lepidoptera: Nymphalidae) during the past 20 million years. *Biological Journal of the Linnean Society* . doi: 10.1111/j.1095-8312.2009.01208.x

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* , *10* (1), 34. doi: 10.1186/1742-9994-10-34

Luo, A., Ling, C., Ho, S. Y. W., & Zhu, C.-D. (2018). Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. *Systematic Biology* , *67* (5), 830–846. doi: 10.1093/sysbio/syy011

Martinet, B., Lecocq, T., Brasero, N., Biella, P., UrbanovÁ, K., ValterovÁ, I., ... Rasmont, P. (2018). Following the cold: geographical differentiation between interglacial refugia and speciation in the arcto-alpine species complex Bombus monticola (Hymenoptera: Apidae). *Systematic Entomology* . doi: 10.1111/syen.12268

McKendrick, L., Provan, J., Fitzpatrick, Ú., Brown, M. J. F., Murray, T. E., Stolle, E., & Paxton, R. J. (2017). Microsatellite analysis supports the existence of three cryptic species within the bumble bee Bombus lucorum sensu lato. *Conservation Genetics* . doi: 10.1007/s10592-017-0965-3

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology* . doi: 10.1371/journal.pbio.0030422

Mutanen, M., Kivelä, S. M., Vos, R. A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., ... Godfray, H. C. J. (2016). Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. *Systematic Biology* , *65* (6), 1024–1040. doi: 10.1093/sysbio/syw044

Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist* . doi: 10.1086/282771

Neumeyer, R., Baur, H., Guex, G. D., & Praz, C. (2014). A new species of the paper wasp genus Polistes (Hymenoptera, Vespidae, Polistinae) in Europe revealed by morphometrics and molecular analyses. *ZooKeys* . doi: 10.3897/zookeys.400.6611

Nichols, H. J., Jordan, N. R., Jamie, G. A., Cant, M. A., & Hoffman, J. I. (2012). Fine-scale spatiotemporal patterns of genetic variation reflect budding dispersal coupled with strong natal philopatry in a cooperatively breeding mammal. *Molecular Ecology* , *21* (21), 5348–5362. doi: 10.1111/mec.12015

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., ... Pevzner, P. A. (2013). *Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads BT - Research in Computational Molecular Biology* (M. Deng, R. Jiang, F. Sun, & X. Zhang, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg.

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

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* . doi: 10.1093/bioinformatics/bty633

Pauly, A., Noël, G., Sonet, G., Notton, D. G., & Boevé, J. L. (2019). Integrative taxonomy resuscitates two species in the lasioglossum villosulum complex (Kirby, 1802) (hymenoptera: Apoidea: Halictidae).*European Journal of Taxonomy* . doi: 10.5852/ejt.2019.541

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* . doi: 10.1111/1755-0998.12129

Praz, C., Müller, A., & Genoud, D. (2019). Hidden diversity in European bees: Andrena amieti sp. n., a new Alpine bee species related to Andrena bicolor (Fabricius, 1775) (Hymenoptera, Apoidea, Andrenidae). *Alpine Entomology* , *3* , 11–38. doi: 10.3897/alpento.3.29675

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. doi: 10.1093/sysbio/syy032

Reemer, M., Groenenberg, D. S. J., Van Achterberg, C., & Peeters, T. M. J. (2008). Taxonomic assessment of Andrena rosae and a stragulata by DNA-sequencing (Hymenoptera: Apoidea: Andrenidae). *Entomologia Generalis* .

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evolutionary Biology* . doi: 10.1186/1471-2148-12-196

Schmid-Egger, C., & Scheuchl, E. (1997). llustrierte Bestimmungstabellen der Wildbienen Deutschlands und Österreichs unter Berücksichtigung der Arten der Schweiz. *Monografien Entomologie Hymenoptera* , *0030* , 1–180.

Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., & Hebert, P. D. N. (2015). DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Molecular Ecology Resources* , *15* (4), 985–1000. doi: 10.1111/1755-0998.12363

Soltani, G. G., Bénon, D., Alvarez, N., & Praz, C. J. (2017). When different contact zones tell different stories: Putative ring species in the Megachile concinna species complex (Hymenoptera: Megachilidae). *Biological Journal of the Linnean Society* . doi: 10.1093/biolinnean/blx023

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences* , *105* (36), 13486–13491. doi: 10.1073/pnas.0803076105

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* ,*30* (9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods in Ecology and Evolution* , *4* (12), 1101–1110. doi: 10.1111/2041-210X.12107

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* , *21* (16), 3907–3930. doi: 10.1111/j.1365-294X.2012.05664.x

van der Meer, F., Reemer, M., Peeters, T. M. J., & Neve, A. (2006). De roodzandbij Andrena rosae in de Zuid-Hollandse Biesbosch (Hymenoptera: apoidea: andrenidae). *Nederlandse Faunistische Mededelingen / Nationaal Natuurhistorisch Museum* .

Westrich, P. (2014). Beitrag zur Diskussion über den taxonomischen Status von Andrena rosae Panzer 1801 (Hymenoptera, Apidae. *Eucera* .

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

Wilkinson, L. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics* . doi: 10.1111/j.1541-0420.2011.01616.x

Williams, K. A., Lelej, A. S., & Thaochan, N. (2019). New species of Myrmosinae (Hymenoptera: Mutillidae) from Southeastern Asia. *Zootaxa* , *4656* (3), 525–534. doi: 10.11646/zootaxa.4656.3.9

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences* . doi: 10.1073/pnas.0913022107

Yang, Z., & Rannala, B. (2014). Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Molecular Biology and Evolution* , *31* (12), 3125–3135. doi: 10.1093/molbev/msu279

Zarza, E., Connors, E. M., Maley, J. M., Tsai, W. L. E., Heimes, P., Kaplan, M., & McCormack, J. E. (2018). Combining ultraconserved elements and mtDNA data to uncover lineage diversity in a Mexican highland frog (Sarcohyla ; Hylidae). *PeerJ* , *6* , e6045. doi: 10.7717/peerj.6045

Data accessibility

Final concatenated fasta files per species complex are available at Dryad (**Currently curating files for upload**). COI sequences are accessible on BOLD through the following DOIs: **Currently curating files for upload**

Authors contribution

Sampling was mainly performed by C.P. Species morphological identifications were realized by C.P. Laboratory, bioinformatic and statitical work was mainly executed by M.G. Manuscript was written by M.G., C.P. and J.F.

Tables

Table 1: Analysis of molecular variance (AMOVA) performed per species complex. AMOVAs were performed on the UCEs dataset using the dartR and pegas R packages. Statistical significant tests (p-values < 0.0001) are highlighted in bold followed an asterisks (*). Significance was assessed through 10,000 permutations. For species complexes composed of more than two species, the analyses was run twice, once with all species and once with only the closed phylogenetically species.

Species complex	**Variance source**	**df**	**SSD**	**MSD**	**% variance**
A. amieti mtDNA ML1 / ML2	Between lineages	1	0.002	0.002	1.75%
	Error	26	0.047	0.002	98.25%
	Total	27	0.049	0.002	100%
A. bicolor mtDNA ML1 / ML2	Between lineages	1	0.026	0.026	**43.43%***
	Error	53	0.061	0.001	56.57%
	Total	54	0.087	0.002	100%
A. barbareae / cineraria	Between lineages	1	0.016	0.016	**52.74%***
	Error	10	0.021	0.002	47.26%
	Total	11	0.037	0.003	100%
A. carantonica / trimmerana / rosae	Between lineages	2	0.145	0.072	**72.97%***
	Error	23	0.069	0.003	27.03%
	Total	25	0.214	0.009	100%

Species complex	**Variance source**	**df**	**SSD**	**MSD**	**% variance**
A. carantonica / trimmerana	Between lineages	1	0.008	0.008	**27.03%***
	Error	17	0.031	0.002	72.97%
	Total	18	0.038	0.002	100%
A. dorsata / propinqua	Between lineages	1	0.012	0.012	**39.33%***
	Error	17	0.026	0.002	60.67%
	Total	18	0.039	0.003	100%
L. alpigenum / bavaricum / cupromicans	Between lineages	2	0.221	0.110	**77.03%***
	Error	16	0.081	0.005	22.97%
	Total	18	0.302	0.017	100%
L. bavaricum / cupromicans	Between lineages	1	0.028	0.028	**38.28%***
	Error	9	0.057	0.006	61.73%
	Total	10	0.085	0.009	100%
N. goodeniana / succincta	Between lineages	1	0.030	0.030	**66.13%***
	Error	13	0.025	0.002	33.87%
	Total	14	0.055	0.004	100%

Table 2: Pairwise Fst and Tajima-Nei's net genetic distance per species complex. Statistical significance (p [?] 0.05) of pairwise Fst were assessed through 10,000 bootstraps and are depicted by asterisks (*). Only the Fst computed between both mitochondrial lineages (ML) of*A. amieti* was not significant (highlighted in bold). Fst tests were computed on the SNP matrixes in R using the dartR package; Tajima-Nei's net genetic distances were computed on the concatenated fasta files using MEGA-x.

Tajima-Nei's D \ Fst	A. allosa	A. amieti **ML1**	A. amieti **ML1**	A. amieti **ML2**	A. amieti **ML2**	A. bicolor **ML1**	A. bicolor **ML1**	*A. bicolor* **ML2**	*A. bicolor* **ML2**	*A.* **sp3**	*A.* **sp3**	*A* n
A. allosa	-	0.00261	0.00261	0.00299	0.00299	0.01080	0.01080	0.01117	0.01117	0.01179	0.01179	0
A. amieti **ML1**	0.196*	-	-	0.00027	0.00027	0.00923	0.00923	0.00961	0.00961	0.00962	0.00962	0
A. amieti **ML2**	0.203*	**- 0.031**	**- 0.031**	-	-	0.01048	0.01048	0.01085	0.01085	0.01145	0.01145	0
A. bicolor **ML1**	0.340*	0.313*	0.313*	0.390*	0.390*	-	-	0.00231	0.00231	0.00510	0.00510	0

Tajima-Nei's D \ Fst	A. allosa	A. amieti **ML1**	A. amieti **ML1**	A. amieti **ML2**	A. amieti **ML2**	A. bicolor **ML1**	A. bicolor **ML1**	*A. bicolor* **ML2**	*A. bicolor* **ML2**	*A.* **sp3**	*A.* **sp3**	A
A. bicolor **ML2**	0.482*	0.456*	0.456*	0.533*	0.533*	0.138*	0.138*	-	-	0.00549	0.00549	0
A. **sp3**	0.62*	0.401*	0.401*	0.637*	0.637*	0.294*	0.294*	0.342*	0.342*	-	-	0
A. montana	0.878	0.773*	0.773*	0.841*	0.841*	0.629*	0.629*	0.654*	0.654*	0.782*	0.782*	-
	A. barbareae	A. barbareae	A. barbareae	A. barbareae	A. cineraria	A. cineraria	A. cineraria	A. cineraria	A. vaga	A. vaga	A. vaga	A
A. barbareae	-	-	-	-	0.00294	0.00294	0.00294	0.00294	0.01022	0.01022	0.01022	0
A. cineraria	0.165*	0.165*	0.165*	0.165*	-	-	-	-	0.01065	0.01065	0.01065	0
A. vaga	0.293*	0.293*	0.293*	0.293*	0.330*	0.330*	0.330*	0.330*	-	-	-	-
	A. carantonica	A. carantonica	A. carantonica	A. carantonica	A. trimmerana	A. trimmerana	A. trimmerana	A. trimmerana	A. rosae	A. rosae	A. rosae	A
A. carantonica	-	-	-	-	0.00061	0.00061	0.00061	0.00061	0.00672	0.00672	0.00672	0
A. trimmerana	0.088*	0.088*	0.088*	0.088*	-	-	-	-	0.00617	0.00617	0.00617	0
A. rosae	0.355*	0.355*	0.355*	0.355*	0.312*	0.312*	0.312*	0.312*	-	-	-	-
	A. dorsata	A. dorsata	A. dorsata	A. dorsata	A. propinqua	A. propinqua	A. propinqua	A. propinqua	A. congruens	A. congruens	A. congruens	A
A. dorsata	-	-	-	-	0.00134	0.00134	0.00134	0.00134	0.01439	0.01439	0.01439	0
A. propinqua	0.142*	0.142*	0.142*	0.142*	-	-	-	-	0.01423	0.01423	0.01423	0
A. congruens	0.795*	0.795*	0.795*	0.795*	0.826*	0.826*	0.826*	0.826*	-	-	-	-
	L. alpigenum	L. alpigenum	L. bavaricum	L. bavaricum	L. bavaricum	L. bavaricum	L. cupromicans	L. cupromicans	L. cupromicans	L. cupromicans	L. nitidulum	L
L. alpigenum	-	-	0.00406	0.00406	0.00406	0.00406	0.00418	0.00418	0.00418	0.00418	0.01134	0
L. bavaricum	0.468*	0.468*	-	-	-	-	0.00083	0.00083	0.00083	0.00083	0.01114	0
L. cupromicans	0.464*	0.464*	- 0.002	- 0.002	- 0.002	- 0.002	-	-	-	-	0.01112	0
L. nitidulum	0.442*	0.442*	0.140*	0.140*	0.140*	0.140*	0.057*	0.057*	0.057*	0.057*	-	-
	N. goodeniana	N. goodeniana	N. goodeniana	N. goodeniana	N. succincta	N. succincta	N. succincta	N. succincta	N. bifasciata	N. bifasciata	N. bifasciata	N
N. goodeniana	-	-	-	-	0.00381	0.00381	0.00381	0.00381	0.01883	0.01883	0.01883	0

Tajima-Nei's D \ Fst	A. allosa	A. amieti **ML1**	A. amieti **ML1**	A. amieti **ML2**	A. amieti **ML2**	A. bicolor **ML1**	A. bicolor **ML1**	*A. bicolor* **ML2**	*A. bicolor* **ML2**	*A.* **sp3**	*A.* **sp3**	A
N. succincta	0.251*	0.251*	0.251*	0.251*	-	-	-	-	0.02014	0.02014	0.02014	0
N. bifasciata	0.490*	0.490*	0.490*	0.490*	0.548*	0.548*	0.548*	0.548*	-	-	-	-

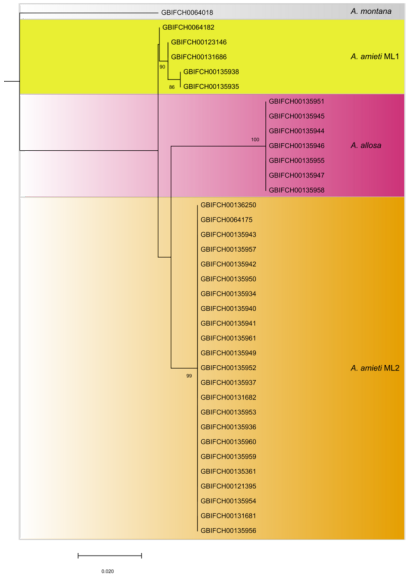
Table3. Summary table showing number of clusters by methods. Results for each method are provides either as main Figure (Figure 1; RAxML) or as Supplementary information (S8-S10). For each species groups, methods providing identical results than morphological identifications are highlighted in bold.

A. amieti mtDNA **ML1** / **ML2**
A. bicolor mtDNA **ML1** / **ML2**
A. barbareae / cineraria
A. carantonica / trimmerana / rosae
A. dorsata / propinqua
L. alpigenum / bavaricum / cupromicans
N. goodeniana / succincta
Number of clusters for the optimal K solution (i.e. lowest BIC value). Alternative k solutions with similar BIC values than

Figures

Figure 1: Maximum likelihood phylogenetic trees obtained for COI and UCEs datasets. Trees were built with RAxML v8.2.11 using the GTR gamma model using 100 bootstrap replicates. Only bootstrap probability higher than 70% are shown.

RAxML COI



RAxML UCEs