Received: 21 February 2022 | Accepted: 10 May 2022

DOI: 10.1111/syen.12557

ORIGINAL ARTICLE

Systematic Entomology

Royal Entomological Society

Delimiting continuity: Comparison of target enrichment and double digest restriction-site associated DNA sequencing for delineating admixing parapatric *Melitaea* butterflies

Mukta Joshi[1] | Marianne Espeland[2] | Vlad Dincă[1] | Roger Vila[3] | Mohadeseh S. Tahami[1] | Lars Dietz[2] | Christoph Mayer[2] | Sebastian Martin[2] | Leonardo Dapporto[4] | Marko Mutanen[1]

[1]Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

[2]Leibniz Institute for the Analysis of Biodiversity Change (LIB), Zoological Research Museum Alexander Koenig, Bonn, Germany

[3]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

[4]ZEN Lab, Dipartimento di Biologia, dell' Università di Firenze, Sesto Fiorentino, Italy

Correspondence
Mukta Joshi, Ecology and Genetics Research Unit, University of Oulu, Pentti Kaiteran katu 1, 90570 Oulu, Finland.
Email: mukta.joshi@oulu.fi

Funding information
Academy of Finland, Grant/Award Numbers: 314702, 328895; MCIN/AEI, Grant/Award Number: PID2019-107078GB-I00

Abstract

Parapatrically distributed taxa pose a challenge for species delimitation due to the presence of gene flow and inherent arbitrariness of exactly defining the species boundaries in such systems. We tackled the problem of species delimitation in a parapatric species pair of *Melitaea* butterflies using two popular genomic methods—double digest restriction-site associated DNA sequencing (ddRAD) and target enrichment. We compared newly generated target enrichment dataset with 1733 loci to the already available ddRAD data from a previous study on the same set of specimens using a suite of phylogenetic, population genetic, and species delimitation methods. We recovered consistent phylogenetic relationships across the datasets, both demonstrating the presence of a genetically distinct Balkan lineage and paraphyly of *Melitaea athalia* with respect to *Melitaea celadussa*. Population genetic STRUCTURE analyses supported the presence of two species when using ddRAD data, but three species when using target enrichment, while a Bayes factor delimitation analysis found both two and three species scenarios equally decisive in both datasets. As the results obtained from both methods were largely congruent, we discuss some practical considerations and benefits of target enrichment over RAD sequencing. We conclude that the choice of method of genomic data collection does not influence the results of phylogenetic analyses at alpha taxonomic level, given a sufficient number of loci. Finally, we recommend a solution for delineating species in parapatric scenarios by proposing that parapatric taxa be consistently classified as subspecies or complete species, but not both, to promote taxonomic stability.

KEYWORDS
admixture, ddRAD sequencing, genomics, hybrid zone, parapatry, species delimitation, target enrichment

INTRODUCTION

Most biologists would likely agree that, of all the taxonomic units, that of a species is fundamental (de Queiroz, 2005; Hohenegger, 2014). The reality of species as a truly discrete biological entity is however unlikely to hold true universally as speciation is usually a long and gradual process (de Queiroz, 2007). The ability to categorize biological diversity into distinct species, that is, species delimitation, is a prerequisite for the identification of individual organisms. The latter, in turn, is important in most biological research, including conservation, but also in many other disciplines and areas of society, such as legislation and food industry (Tewksbury et al., 2014). Despite the global efforts to document and describe the bewildering diversity of life on earth, most species remain undescribed or even undiscovered (Mora et al., 2011). In addition to the complexities caused by the usually low differentiation of recently diverged species, traits such as phenotypic plasticity and morphological crypticity displayed by some species seriously complicate the process of species delimitation (Jarman & Elliott, 2000; Knowlton, 1993). Delimitation is further complicated by several other factors. For instance, many species have a large or patchy distribution and show extensive intraspecific variability in geographical space. Additionally, hybridization and introgression between species have been shown to be much more common than previously expected. It has been estimated that 10% of animal species and at least 25% of all plant species are involved in hybridization (Mallet, 2005). Finally, there is a variety of ways to define the term 'species'. Some species concepts have a focus on temporal changes (i.e., when speciation occurs in time) while others focus on spatial (when geographic isolation between diverging populations) or biological (reproductive isolation) aspects of variability (de Queiroz, 2007), rendering the definition of species even more elusive. Despite these difficulties, categorizing biodiversity into species remains a central commitment of taxonomy.

A wide geographical distribution or a geographically biased taxonomic sampling often complicates the process of species delimitation to a significant degree (Linck et al., 2019; Mutanen et al., 2012). When the distributions of species are large, extensive genetic differences among populations in different areas can be maintained, despite migration (Gavrilets et al., 2000). The same may be true on a smaller scale when dispersal is limited by species' own traits or geographic barriers. In European butterflies, the degree of population differentiation is positively correlated with range size and negatively with traits determining their mobility and level of generalism (Dapporto et al., 2019). Also, variation in environmental conditions can result in selection acting differently in different areas within species ranges (Gavrilets et al., 2000). The geographic relationship between species or populations has traditionally been categorized as either allopatric, parapatric, or sympatric (Hendry et al., 2009) although there is a full continuity between these distributional categories. Parapatry refers to the distributional pattern where the separate ranges of two species have a narrow overlap (Bull, 1991), typically within a hybridization zone (Hewitt, 1988). While such circumstances provide an opportunity to peek into the speciation process through reinforcement, parapatry, in the presence of gene flow, poses a particular challenge for species delimitation, as the end point of the speciation process is often vague and gene flow may be extensive (Hendry et al., 2009). Indeed, the well-documented cases of parapatry demonstrate how utterly complicated species delimitation under such circumstances can be. A well-known example of such a system is that of carrion and hooded crows, *Corvus corone* (Linnaeus) and *Corvus cornix* (Linnaeus), in Europe (Poelstra et al., 2013; Saino et al., 1992; Saino & Villa, 1992). These two taxa are now regarded as two different species (Saino et al., 1992) despite high levels of interbreeding and introgression.

Development of high-throughput sequencing technologies has enabled access to whole genomes or large proportions of genomes that can potentially provide useful information to resolve the patterns of evolution from population to very deep phylogenetic levels (Breinholt et al., 2018; Herrera & Shank, 2016). Two of the popular reduced representation approaches, double digest restriction-site associated DNA sequencing (ddRAD; Peterson et al., 2012) and target capture methods such as ultraconserved elements (UCEs; Faircloth et al., 2012), anchored hybrid enrichment (AHE; Lemmon et al., 2012) and other target enrichment approaches (Gnirke et al., 2009, see also Mayer et al., 2016), are now widely used in systematics research. RADseq approaches have been used particularly to elucidate evolutionary patterns at relatively shallow phylogenetic levels (Wagner et al., 2013), while target capture-based approaches can be used for deep as well as shallow-level phylogenomics (Banker et al., 2020).

The species pair *Melitaea athalia* Rottemburg and *Melitaea celadussa* Fruhstorfer represents a well-documented case of widespread and parapatric taxa in European butterflies with the range of *M. athalia* extending through most of the Europe and Western parts of Asia and *M. celadussa* mainly occurring in southern and southwestern parts of Europe (Van Oorschot & Coutsis, 2014). The two species exhibit a narrow hybrid zone mainly occurring in France, Switzerland, Italy and Austria (Van Oorschot & Coutsis, 2014; Higgins, 1955; Platania et al., 2020). The species are morphologically distinguished based on male genitalia, but intermediate male genital characters have been reported from the contact zone (Beuret, 1933; Van Oorschot & Coutsis, 2014). Genomic analyses done using a ddRAD dataset on this species pair revealed wide admixture in the contact zone and across a part of the range of *M. athalia*, but less so in *M. celadussa*, suggesting a mostly unidirectional introgression (Tahami et al., 2021). Additionally, this analysis revealed a genetically diverging and non-admixed, but morphologically indistinguishable clade in the Balkans that had not been detected previously. Using morphometrics of the male genitalia, the same study reported several specimens with intermediate genital characters in the contact zone.

As parapatric systems are likely to reveal a full spectrum of genetic admixture across different systems, taxonomic delimitation of parapatric taxa is bound to be almost inherently arbitrary. However, despite the conceptual nature of their delimitation, there is no consensus in the taxonomic community on the principles of how such taxa should be taxonomically ranked. We believe that genomic approaches could provide excellent means for higher levels of

TABLE 1 Comparison of number of raw reads from the target enrichment and ddRAD datasets

Specimen ID	BioSample accessions	SRA	Taxon based on COI	Country	Number of target enrichment reads after filtering	Number of RAD reads after filtering
RVcoll07E394	SAMN25488010	SRR17830142	*Melitaea athalia*	Romania	2,072,173	4,673,359
RVcoll08M346	SAMN25488011	SRR17830141	*Melitaea athalia*	Romania	2,001,739	1,180,010
RVcoll10A789	SAMN25488012	SRR17830130	*Melitaea athalia*	Estonia	2,219,329	1,361,958
RVcoll12Z197	SAMN25488013	SRR17830119	*Melitaea athalia*	Sweden	1,991,064	3,428,520
RVcoll13U124	SAMN25488014	SRR17830108	*Melitaea celadussa*	Italy	2,050,917	1,047,162
RVcoll13U296	SAMN25488015	SRR17830097	*Melitaea athalia*	Italy	1,836,742	2,348,944
RVcoll13U438	SAMN25488016	SRR17830086	*Melitaea athalia*	Italy	2,084,530	1,793,939
RVcoll14B773	SAMN25488017	SRR17830085	*Melitaea athalia*	Albania	1,614,654	1,334,062
RVcoll14D059	SAMN25488018	SRR17830084	*Melitaea athalia*	Bulgaria	1,720,383	2,904,422
RVcoll14E853	SAMN25488019	SRR17830083	*Melitaea athalia*	Serbia	1,396,629	1,735,214
RVcoll14E859	SAMN25488020	SRR17830140	*Melitaea athalia*	Serbia	1,605,420	1,336,439
RVcoll14E904	SAMN25488021	SRR17830139	*Melitaea athalia*	Serbia	1,703,860	312,305
RVcoll14F303	SAMN25488022	SRR17830138	*Melitaea athalia*	Serbia	1,885,222	514,133
RVcoll14F407	SAMN25488023	SRR17830137	*Melitaea athalia*	Bulgaria	1,464,925	526,291
RVcoll14F538	SAMN25488024	SRR17830136	*Melitaea athalia*	Greece	1,827,776	1,418,752
RVcoll14F650	SAMN25488025	SRR17830135	*Melitaea athalia*	Greece	1,470,771	800,882
RVcoll14F666	SAMN25488026	SRR17830134	*Melitaea athalia*	Greece	1,710,058	899,141
RVcoll14G434	SAMN25488027	SRR17830133	*Melitaea athalia*	Greece	1,830,124	822,392
RVcoll14V075	SAMN25488028	SRR17830132	*Melitaea athalia*	Ukraine	1,819,607	2,639,743
RVcoll15I360	SAMN25488029	SRR17830131	*Melitaea athalia*	Austria	1,736,020	3,447,314
RVcoll15P033	SAMN25488030	SRR17830129	*Melitaea athalia*	Ukraine	1,945,474	2,394,573
RVcoll16H415	SAMN25488031	SRR17830128	*Melitaea athalia*	Sweden	1,757,794	6,630,426
RVcoll16I052	SAMN25488032	SRR17830127	*Melitaea athalia*	Poland	1,624,217	1,399,738
RVcoll16J000	SAMN25488033	SRR17830126	*Melitaea athalia*	Slovakia	1,821,930	1,669,269
RVcoll16J612	SAMN25488034	SRR17830125	*Melitaea athalia*	Russia	1,927,534	3,477,748
MAT-SG-W-144	SAMN25488035	SRR17830124	*Melitaea celadussa*	Switzerland	1,721,973	2,207,274
RVcoll08J851	SAMN25488036	SRR17830123	*Melitaea celadussa*	Spain	1,726,868	2,139,191
RVcoll08L852	SAMN25488037	SRR17830122	*Melitaea celadussa*	Spain	1,538,438	1,162,161
RVcoll08M074	SAMN25488038	SRR17830121	*Melitaea celadussa*	Spain	1,880,216	1,233,798
RVcoll08M915	SAMN25488039	SRR17830120	*Melitaea celadussa*	Spain	1,735,217	1,319,987
RVcoll08P221	SAMN25488040	SRR17830118	*Melitaea celadussa*	Spain	1,942,107	2,361,835
RVcoll11H561	SAMN25488041	SRR17830117	*Melitaea celadussa*	Italy	1,676,314	912,613
RVcoll11H741	SAMN25488042	SRR17830116	*Melitaea celadussa*	Italy	1,617,340	2,127,974
RVcoll11I507	SAMN25488043	SRR17830115	*Melitaea celadussa*	Spain	1,835,635	1,682,848
RVcoll11I949	SAMN25488044	SRR17830114	*Melitaea celadussa*	France	1,843,398	1,087,215
RVcoll12O623	SAMN25488045	SRR17830113	*Melitaea celadussa*	France	1,365,257	1,844,014
RVcoll12P926	SAMN25488046	SRR17830112	*Melitaea celadussa*	France	2,152,092	1,807,757
RVcoll12Q105	SAMN25488047	SRR17830111	*Melitaea celadussa*	France	1,799,934	1,903,771
RVcoll12Q106	SAMN25488048	SRR17830110	*Melitaea celadussa*	France	1,818,742	782,461
RVcoll13S845	SAMN25488049	SRR17830109	*Melitaea celadussa*	Portugal	1,769,707	4,178,173
RVcoll13U092	SAMN25488050	SRR17830107	*Melitaea celadussa*	Italy	2,030,995	2,365,119
RVcoll14E220	SAMN25488051	SRR17830106	*Melitaea celadussa*	Spain	1,915,377	5,057,299
RVcoll14J820	SAMN25488052	SRR17830105	*Melitaea celadussa*	France	1,838,442	2,919,525
RVcoll14L240	SAMN25488053	SRR17830104	*Melitaea celadussa*	Italy	1,596,947	2,293,724

(Continues)

TABLE 1 (Continued)

Specimen ID	BioSample accessions	SRA	Taxon based on COI	Country	Number of target enrichment reads after filtering	Number of RAD reads after filtering
RVcoll15A654	SAMN25488054	SRR17830103	*Melitaea celadussa*	Italy	2,104,662	684,105
RVcoll15G145	SAMN25488055	SRR17830102	*Melitaea celadussa*	France	1,772,903	1,099,637
RVcoll15G841	SAMN25488056	SRR17830101	*Melitaea celadussa*	Italy	1,952,921	971,078
RVcoll15I495	SAMN25488057	SRR17830100	*Melitaea celadussa*	Austria	1,850,398	1,365,567
RVcoll15L146	SAMN25488058	SRR17830099	*Melitaea celadussa*	Italy	1,741,727	273,881
RVcoll15M133	SAMN25488059	SRR17830098	*Melitaea celadussa*	France	1,625,035	2,962,880
RVcoll15N014	SAMN25488060	SRR17830096	*Melitaea celadussa*	Italy	2,298,510	2,911,962
RVcoll16C754	SAMN25488061	SRR17830095	*Melitaea celadussa*	Italy	1,436,688	818,831
MAT-UR-I-146	SAMN25488062	SRR17830094	*Melitaea celadussa*	Switzerland	1,802,507	1,481,541
MAT-LU-K-122	SAMN25488063	SRR17830093	*Melitaea celadussa*	Switzerland	1,646,303	1,511,226
MAT-SG-W-135	SAMN25488064	SRR17830092	*Melitaea celadussa*	Switzerland	1,887,704	1,630,488
MAT-SG-W-137	SAMN25488065	SRR17830091	*Melitaea celadussa*	Switzerland	1,740,049	863,860
MAT-SG-W-138	SAMN25488066	SRR17830090	*Melitaea celadussa*	Switzerland	1,742,294	1,308,809
MAT-SG-W-140	SAMN25488067	SRR17830089	*Melitaea celadussa*	Switzerland	1,661,150	489,925
ZFMK-TIS-8000434	SAMN25488068	SRR17830088	*Melitaea caucasogenita*	Georgia	1,760,707	2,034,813
RVcoll14V076	SAMN25488069	SRR17830087	*Melitaea britomartis*	Ukraine	1,466,749	1,953,126

standardization for delimiting parapatric taxa by enabling the evaluation of important biological properties, such as levels of introgression and genetic divergence, in a quantitative way and based on large amounts of standardized genetic markers (Dietz et al., 2019; Eberle et al., 2020). Hence, motivated by the low number of studies that use genomics to assess parapatric species systems, we tackled the question of their delimitation based on two powerful genomic approaches, ddRAD sequencing and target enrichment. We compared the two methods making use of phylogenetic, population genetic and species delimitation methods, because species delimitation essentially acts at the interface of phylogenetic and population genetic levels. Apart from comparison of the results and usefulness of ddRAD sequencing versus target enrichment, we provide insights into the conceptual aspects of the delimitation of parapatric taxa with admixture in order to advance the stability and consistency of their delimitation.

MATERIALS AND METHODS

Sample preparation

Samples were collected from various locations across Europe and were representative for the range of the targeted taxa, including their contact zone (Table 1). Initially, a set of 81 specimens was used for ddRAD analyses (Tahami et al., 2021) and a subset of 60 specimens was selected for target enrichment analyses. These included specimens of *M. athalia* and *M. celadussa*, as well as one specimen putatively attributed to *Melitaea caucasogenita* Verity and one *Melitaea britomartis* Assmann as outgroup taxa. Because some specimens were females and it was not clear how well they can be identified based on genitalia, DNA barcodes were used for a priori identification and labelling of specimens. We also categorized specimens occurring in the supposed contact zone using the biodecrypt approach (Platania et al., 2020), which implements dedicated functions to identify areas of sympatry among cryptic taxa based on a subset of identified specimens (details are given in the Supporting information). As per this analysis, we inferred 11 specimens as belonging to the contact zone (Table S2). Genomic DNA was extracted from two thirds of the thorax from either ethanol preserved or dry specimens. DNA extraction was done using the QIAGEN DNeasy Blood and Tissue kit following the protocol provided by the manufacturer. DNA extracts were visualized on 1% agarose gels.

Target enrichment laboratory procedures

Target enrichment bait design followed Mayer et al. (2021) where the final probe kit targets 2953 CDS regions in 1753 nuclear genes. This kit was developed with BaitFisher version 1.2.8 (Mayer et al., 2016) and is referred to as the LepZFMK 1.0 kit. The DNA concentration of each sample was quantified using a Promega Quantus fluorometer, and the initial fragment lengths were measured with a Fragment Analyser (Agilent Technologies Inc.). About 100 ng absolute amount of DNA was taken for further processing as per standard Agilent protocol. The genomic DNA was subjected to random mechanical shearing to an average size between 250 and 300 bp followed by an end-repair reaction and ligation of adenine residue to the 3′ end of the blunt fragments (A-tailing) to allow ligation of barcoded adaptors using the

Agilent SureSelect XT2 Library prep kit. Adaptor ligation and indexing was performed using the New England Biolabs reaction kit to enable dual indexed libraries. The resulting libraries were then PCR amplified. The number of PCR cycles at this stage was determined on the basis of library concentration after A-tailing and a total of eight cycles were used for all samples. The concentration of PCR-amplified libraries was checked using the Quantus fluorometer and fragment distribution was analysed using the Fragment Analyzer. After the library construction, custom SureSelect baits (Agilent Technologies, SureSelect Custom Baits size 6–11.9 Mb) were used for solution-based target enrichment of a pool containing eight libraries. The hybridization was performed on each pool with SureSelect XT2 precapture ILM module following the manufacturer's instructions. Enriched libraries were then captured with Streptavidin beads (MyOne Streptavidin T1). These captured libraries were PCR amplified and the final concentration of each captured library was again checked on the Quantus fluorometer. Following the enrichment, pooled libraries were sequenced using Illumina Nextseq 500 mid output to generate paired-end 150-bp reads at Starseq GmbH (Mainz, Germany).

Target enrichment bioinformatics

Adaptor sequences and low-quality regions were trimmed from demultiplexed data with fastq-mcf (Aronesty, 2011). The reads were mapped against the reference gene alignments of the bait regions generated during bait design with the BWA-MEM algorithm in bwa 0.7.17 (available from bio-bwa.sourceforge.net) with the minimum seed length set to 30. Reads for which the mapping was successful were then extracted from the resulting SAM file using a custom Perl script (Dietz et al., 2019) and mapped against a full coding sequence with BWA as described above. Diploid consensus sequences of the regions matching the reference were generated with samtools 1.6 (Li et al., 2009) and bcftools 1.6 (https://github.com/samtools/bcftools). The consensus sequences were further converted into alignments by another custom script (Dietz et al., 2019) to generate the individual gene alignments. This was followed by a gap removal from these alignments. In order to minimize the amount of missing data, which can potentially introduce biases, we generated two datasets: one with the loci that are present in at least 50% of the specimens (TE30) and another with loci that are present in all the specimens (TE60). These alignments were then manually checked in Geneious version 6.1.8 to make sure that there were no misalignments. Finally, individual loci alignments were concatenated for the initial phylogenetic analyses using FASconCAT-G (Kück & Longo, 2014).

Laboratory procedures and bioinformatics for ddRADseq followed Tahami et al. (2021).

SNP calling

We extracted SNPs from the sorted BAM files generated during mapping in the TE30 dataset. SNPs were called using the samtools *mpileup* option piped together with the bcftools *call* option. Then filtering was done using bcftools, keeping one randomly chosen SNP per locus, filtering out indels and multiallelic SNPs. The resulting dataset contained a total of 3164 SNPs, which were used for further analysis.

For ddRAD data, 18,383 SNPs were generated after the assembly and filtering where only those SNP sites for which data was present for at least 20 samples (*m*20) were taken into consideration.

Phylogenetic analyses

For the target enrichment data, the best partitioning scheme for the concatenated dataset was found using IQTREE version 2.0.3 (Chernomor et al., 2016; Minh, Schmidt, et al., 2020b) with option-m TESTMERGEONLY to resemble PartitionFinder and rcluster algorithm (Lanfear et al., 2014), with the rcluster percentage set to 10, under the AICc criterion. The best partitioning scheme was then used as an input to set up a partitioned analysis in IQ-TREE. We used the ultrafast bootstrap approximation with 1000 replicates (Hoang et al., 2018). We also performed a SH-like approximate likelihood ratio test (Guindon et al., 2010) with 1000 bootstrap replicates using the -alrt option. To further reduce the risk of overestimating branch supports, the -bnni option was used. For ddRAD data, the TVM+F+I+G4 model of sequence evolution was used to reconstruct the ML tree in IQTREE using the ultrafast bootstrap approximation to compute 1000 replicates (Tahami et al., 2021). For each of the TE30, TE60, and ddRAD datasets, the NNI searches were performed using 20 best initial trees. The resulting phylogenetic trees were visualized using FigTree (https://github.com/rambaut/figtree/releases) and rooted on *M. britomartis*. We compared the resulting trees using the *tipdiff* function in the R package treespace (Jombart et al., 2017). This function finds the number of differences in ancestry of the tips of two different trees with the same tip labels.

Species tree analyses

Taking into consideration the gene tree-species tree discordance and incomplete lineage sorting that is common in datasets with multiple loci, we inferred a species tree using the summary-based coalescent method ASTRAL-III v. 5.7.3 (Zhang et al., 2018), which is statistically consistent under the multispecies coalescent framework. Model selection on each individual gene alignment from the TE60 dataset was performed using ModelFinder (Kalyaanamoorthy et al., 2017) and tree inference was done in IQTREE. The resulting output gene trees were used as an input for ASTRAL, which generated a species tree along with the quartet score. This score is the fraction of the induced quartet trees in the input set that occur in the species tree. The branch lengths of the ASTRAL species tree are in coalescent units and support values are given as local posterior probabilities (Zhang et al., 2018).

Gene and site concordance factor

We calculated gene and site concordance factors (gCF and sCF) in IQTREE version 2.0.3 (Minh, Hahn, & Lanfear, 2020a) using the same set of gene trees used as an input for ASTRAL. gCF is calculated for each branch of a species tree as the fraction of decisive gene trees concordant with this branch and sCF as a fraction of decisive alignment sites supporting that branch (Minh, Hahn, & Lanfear, 2020a). The gene trees were compared against the species tree generated in IQTREE using a concatenation approach and various metrics were generated for the TE60 dataset. To better understand how concordance factors relate to each other and to the bootstrap values, resulting gCF and sCF values were plotted in R v. 4.0.3 (Lanfear, 2018).

Population genetics

To get an initial idea of genetic clusters of related species present in the dataset, a principal component analysis (PCA) was carried out on the target enrichment SNP dataset and on all genotypes as well as unlinked SNPs obtained from ddRAD dataset using the *dudi.pca* function from the R package *adegenet* (Jombart & Ahmed, 2011). We also performed a STRUCTURE analysis (Pritchard et al., 2000) to compare the admixture patterns inferred from ddRAD data and the target enrichment data. For this, our target enrichment SNP data in vcf format was converted to structure format (.str) using PGDSpider version 2.1.1.5 (Lischer & Excoffier, 2012). We tested five putative numbers of clusters, $K = 1$–5, with 10 iterations for each K. To determine the optimal number of genetic clusters (K), we used the ΔK method in STRUCTURE HARVESTER (Earl & vonHoldt, 2012; Evanno et al., 2005) with 500,000 generations for the Markov chain and a value of 100,000 as burn-in. The same parameters were used for the target enrichment and the ddRAD dataset. We then aligned the cluster assignments of $K = 2$, $K = 3$, and $K = 4$ across all the 10 replicates in CLUMPP (Jakobsson & Rosenberg, 2007) and used DISTRUCT (Rosenberg, 2004) to visualize the patterns of admixture from aligned clusters for both the datasets.

Species delimitation

We performed a coalescent-based Bayes factor species delimitation implemented in SNAPP (Bryant et al., 2012) using the BFD* method (Leaché et al., 2014) on both target enrichment and ddRAD datasets. For ddRAD data, we used unlinked SNPs to test the species delimitation scenarios in SNAPP. We generated a subset of 30 taxa (Table S1) and tested two alternate scenarios. In the first scenario, the Balkan lineage of *M. athalia* was separated from the rest of the *M. athalia* and *M. celadussa* populations (Run C, Table 3) and in the second scenario *M. athalia* (except for Balkan lineage of *M. athalia*) was lumped together with *M. celadussa* (Run D, Table 3). The current taxonomy comprises two species—*M. athalia* (also including the Balkan lineage) and *M. celadussa*. For each scenario, a path-sampling analysis was carried out with 10 million generations. The run parameters including priors are described in Supporting information. After the run for each of the alternate scenario and current taxonomy, SNAPP generated Marginal likelihood estimates (MLE). Bayes factors (Kass & Raftery, 1995) were calculated as the difference between current taxonomy and each of the alternate scenarios, which is then used to assess the strength of species delimitation models. As per the BF scale, $0 < BF < 2$ is not worth more than a bare mention, $2 < BF < 6$ indicates positive evidence, $6 < BF < 10$ represents strong support, $BF > 10$ indicates decisive support. The calculation of Bayes Factors and ranking of different models followed the steps mentioned in Leaché and Bouckaert (2018).

In addition to the SNAPP analysis, we also tested the alternate species assignments using the *tr2* program, a multilocus species delimitation method that finds the best delimitation based on a distribution model of rooted triplets (Fujisawa et al., 2016). We used rooted individual gene trees from the TE60 dataset as input. The two models we tested, which correspond to the two-species and three-species hypotheses respectively, were compared against a null model (which assumes the presence of a single species) based on -log(likelihood) scores.

RESULTS

Overview of the datasets

After the read filtering step, an average of 1.7 million reads was recovered across all the specimens for the target enrichment dataset (Table 1). For the TE30 dataset, the average number of informative loci retained was 1733 (SD = 21.7, Table 2), with an average amount of missing data of 14.15% (SD = 0.025, Table 2). From this dataset, we further removed the loci with zero or very few variable sites (number of variable sites = 0, 1 or 2). The final average number of loci retained was 1578. For the TE60 dataset, 1031 loci were retained with the average amount of missing data dropping to 0.60% (Table 2). After removing the loci with zero or few variable sites, 1002 loci were retained. For the ddRAD dataset, the average number of loci retained after assembly was 5071 (Table 2), but the percentage of missing data was much higher than for the target enrichment dataset (average 78.9%, Table 2).

Target enrichment phylogenetics

Initial identification and labelling of our specimens followed the information provided by their mitochondrial COI barcodes. Two separate trees were generated for the two target enrichment datasets, TE30 and TE60. In both cases, six specimens of *M. athalia* originating from the Balkans, mainly the countries Albania, Bulgaria, Serbia and Greece (Balkan clade, atha14F407, atha14F660, atha14F666, atha14B773, atha14E859, atha14E853) were recovered as a distinct lineage that was sister to the rest with both ultrafast bootstrap and SH-aLRT support values of 100% (Figures 1 and 3a). However, this Balkan clade did not include all the specimens from the Balkan region that we

TABLE 2 Overview of number of loci and missing data for the target enrichment and ddRAD datasets

Specimen ID	Number of loci recovered (TE30 dataset)	Missing data TE30	Number of loci recovered (TE60 dataset)	Missing data TE60	Number of ddRAD loci after assembly	% missing data ddRAD
RVcoll07E394	1745	11.25%	1031	0.77%	4180	82.35%
RVcoll08M346	1734	11.76%	1031	0.70%	5737	77.40%
RVcoll10A789	1772	10.51%	1031	0.74%	5075	78.87%
RVcoll12Z197	1746	12.02%	1031	0.68%	7994	70.82%
RVcoll13U124	1749	11.20%	1031	0.70%	4658	79.86%
RVcoll13U296	1738	12.62%	1031	0.71%	5609	75.94%
RVcoll13U438	1754	12.89%	1031	0.76%	3557	84.71%
RVcoll14B773	1708	18.44%	1031	0.32%	5597	74.07%
RVcoll14D059	1736	15.14%	1031	0.64%	5568	77.59%
RVcoll14E853	1650	20.42%	1031	0.41%	4763	76.31%
RVcoll14E859	1693	17.66%	1031	0.36%	4374	77.78%
RVcoll14E904	1732	14.51%	1031	0.70%	3538	84.74%
RVcoll14F303	1752	14.67%	1031	0.60%	3875	82.87%
RVcoll14F407	1674	19.65%	1031	0.34%	3256	82.57%
RVcoll14F538	1739	13.71%	1031	0.64%	4439	81.97%
RVcoll14F650	1680	17.50%	1031	0.32%	3939	78.96%
RVcoll14F666	1705	17.21%	1031	0.25%	3352	82.68%
RVcoll14G434	1732	13.75%	1031	0.66%	5943	76.44%
RVcoll14V075	1747	15.63%	1031	0.82%	7873	69.87%
RVcoll15I360	1739	16.75%	1031	0.83%	4106	81.69%
RVcoll15P033	1746	12.89%	1031	0.72%	7043	72.49%
RVcoll16H415	1737	13.92%	1031	0.62%	6500	73.75%
RVcoll16I052	1719	15.55%	1031	0.71%	5416	77.71%
RVcoll16J000	1738	13.80%	1031	0.73%	5315	77.81%
RVcoll16J612	1749	10.99%	1031	0.72%	5091	78.99%
MAT-SG-W-144	1737	14.77%	1031	0.68%	9051	69.94%
RVcoll08J851	1747	13.70%	1031	0.53%	5194	77.52%
RVcoll08L852	1718	15.97%	1031	0.39%	5205	77.86%
RVcoll08M074	1752	12.25%	1031	0.33%	4903	78.12%
RVcoll08M915	1731	12.97%	1031	0.58%	6346	74.06%
RVcoll08P221	1743	11.11%	1031	0.58%	6765	72.63%
RVcoll11H561	1738	14.12%	1031	0.42%	3955	82.87%
RVcoll11H741	1711	15.14%	1031	0.42%	4558	81.04%
RVcoll11I507	1739	13.60%	1031	0.16%	4070	82.26%
RVcoll11I949	1743	12.64%	1031	0.66%	5707	75.05%
RVcoll12O623	1717	18.50%	1031	0.59%	4894	78.54%
RVcoll12P926	1754	9.95%	1031	0.66%	4403	80.81%
RVcoll12Q105	1742	13.55%	1031	0.63%	4940	78.32%
RVcoll12Q106	1756	13.40%	1031	0.65%	4191	82.34%
RVcoll13S845	1725	14.11%	1031	0.30%	3919	82.75%
RVcoll13U092	1754	11.37%	1031	0.63%	3899	83.56%
RVcoll14E220	1745	11.85%	1031	0.18%	2454	88.80%
RVcoll14J820	1756	13.04%	1031	0.72%	6539	72.97%
RVcoll14L240	1754	16.59%	1031	0.71%	5047	78.37%

(Continues)

TABLE 2 (Continued)

Specimen ID	Number of loci recovered (TE30 dataset)	Missing data TE30	Number of loci recovered (TE60 dataset)	Missing data TE60	Number of ddRAD loci after assembly	% missing data ddRAD
RVcoll15A654	1763	11.68%	1031	0.63%	2336	89.43%
RVcoll15G145	1733	14.78%	1031	0.69%	4704	78.54%
RVcoll15G841	1745	13.24%	1031	0.65%	6134	75.37%
RVcoll15I495	1739	14.79%	1031	0.71%	4710	79.47%
RVcoll15L146	1715	12.21%	1031	0.65%	3820	83.44%
RVcoll15M133	1729	15.45%	1031	0.67%	5808	74.75%
RVcoll15N014	1750	9.62%	1031	0.59%	5083	79.96%
RVcoll16C754	1727	16.71%	1031	0.67%	5107	78.58%
MAT-UR-I-146	1734	12.91%	1031	0.66%	5116	78.33%
MAT-LU-K-122	1709	15.13%	1031	0.73%	7416	73.32%
MAT-SG-W-135	1747	10.81%	1031	0.72%	8032	73.64%
MAT-SG-W-137	1734	11.78%	1031	0.72%	5749	76.80%
MAT-SG-W-138	1728	14.92%	1031	0.72%	3914	83.14%
MAT-SG-W-140	1729	16.27%	1031	0.72%	3800	83.42%
ZFMK-TIS-8000434	1726	16.13%	1031	0.33%	4745	83.27%
RVcoll14V076	1707	19.24%	1031	0.58%	4978	85.85%

analysed. The remaining specimens of *M. athalia* were found to be paraphyletic with respect to *M. celadussa* (SH-aLRT = 100%, UFBoot = 100%, Figure 1). Within this group, another set of six *M. athalia* individuals from Balkans (atha14V075, atha08M346, atha14F303, atha14F538, atha14D059, and atha14G434) formed a clade sister to the rest (SH-aLRT = 99.8%, UFBoot = 98% in TE60, SH-aLRT = 99.9%, UFBoot = 100% in TE30). We did not find a clear separation of specimens from the contact zone or of specimens with intermediate genital characters (from morphometric analyses done by Tahami et al., 2021).

ASTRAL species tree and concordance factors

The general patterns found in the ASTRAL species tree were congruent with the maximum likelihood (ML) phylogenetic analyses in IQ-TREE, with the Balkan *M. athalia* clade being supported with local posterior probability of 1, and a clade containing paraphyletic *M. athalia* and *M. celadussa* having local posterior probability of 0.93 (Figure 2a). The same set of six *M. athalia* individuals as in the ML analyses grouped together at one end of this clade (posterior probability 0.66). The scoring of this tree gave a final quartet score of 235,095,870, and only 48.11% of all the quartet trees could be found in the species tree (normalized quartet score of 0.48) indicating high gene tree discordance.

In the species tree obtained from the IQ-TREE concatenation approach (Figure S1), higher values for both concordance factors were mainly observed for the Balkan *M. athalia* clade. For the paraphyletic *M. athalia-M. celadussa* grade, and branches within it, gCF values were lower even though sCF values were (moderately) on the higher side. The same was observed when plotting sCF versus gCF (Figure 2b), where most of the branches fall on the lower side for gCF values (near 0 and between 0 and 25) and moderately high side for sCF values (between 25 and 75), while a few branches had higher values of both concordance factors (gCF above 25 and sCF above 75).

Comparison – phylogenetics

We compared a phylogenetic tree obtained from the TE60 dataset with a phylogenetic tree generated using ddRAD data, excluding the taxa that are not present in the target enrichment dataset (Figure 3a). The general phylogenetic relationships were the same in both datasets. Both methods suggested the presence of a distinct Balkan lineage, which in both cases included the same six samples (Figure 3a, individuals indicated in red colour). The rest of the individuals of *M. athalia* were found to be paraphyletic with respect to *M. celadussa*, with the exception of a clade of another set of six *M. athalia* individuals from the Balkans. This clade was observed to be sister to the paraphyletic *M. athalia-M. celadussa* grade and had rather low support values (SH-aLRT = 51.7%, UFBoot = 66%) in the ddRAD analyses, and higher support in the target enrichment analyses (SH-aLRT = 99.8%, UFBoot = 98%). In both cases, the individuals are identified by biodecrypt as belonging to the contact zone formed part of the paraphyletic *M. athalia-M. celadussa* grade. The comparison using *tipdiff* (Figure S2) shows that the number of differences between the two trees for the six diverging Balkan samples is zero, whereas there are two to four differences in the *M. athalia* clade that groups as sister to the paraphyletic grade, and 42–43 differences for the rest of the taxa from the paraphyletic *M. athalia-M. celadussa* grade.

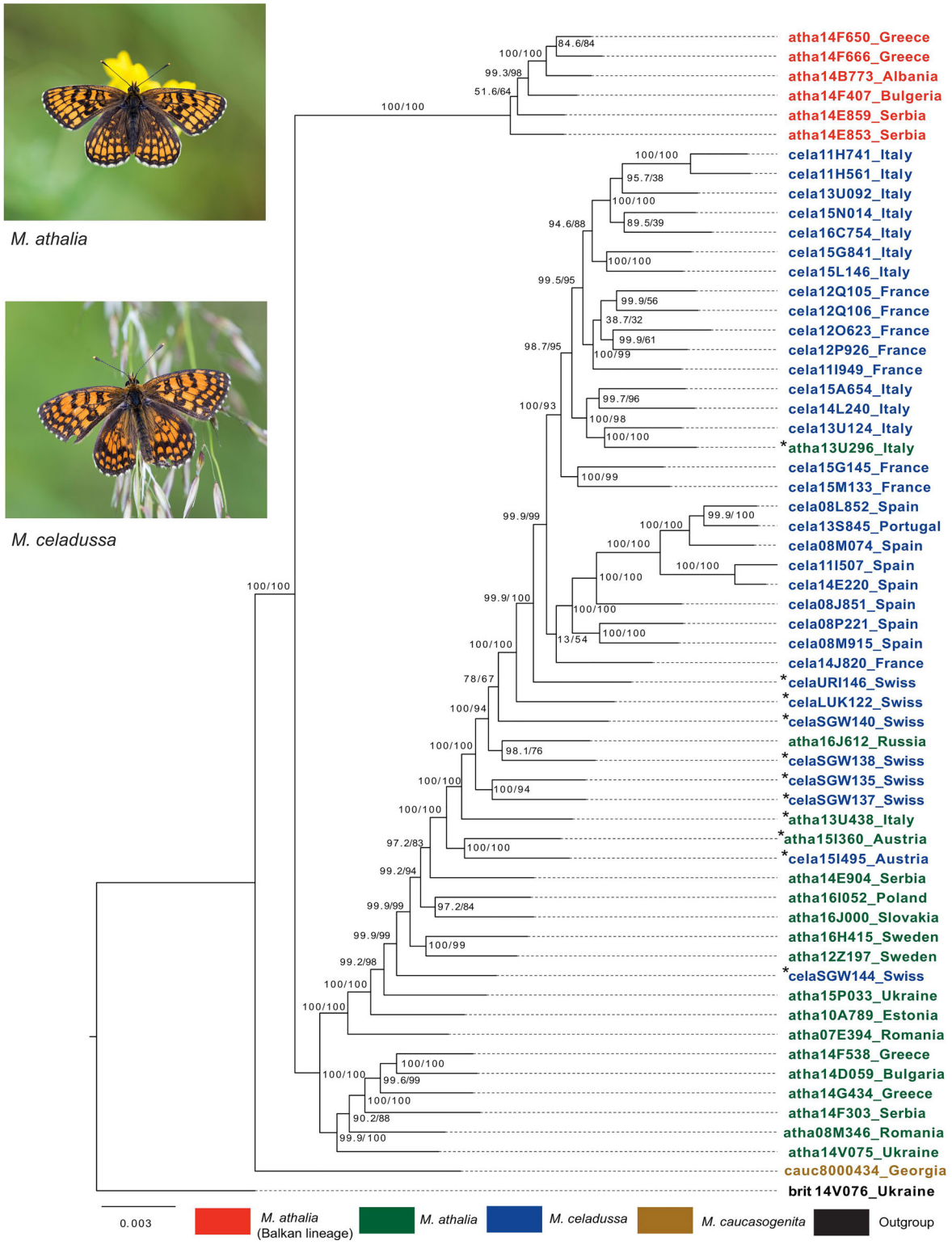

FIGURE 1 Maximum likelihood (ML) phylogenetic tree based on the target enrichment TE30 dataset rooted on *Melitaea britomartis*. Numbers on the branches indicate SH-aLRT/UFBoot support values calculated based on 1000 replicates. Contact zone specimens are indicated with the asterisks.

Comparison – population genetics

The PCA plots for both target enrichment and ddRAD showed a clade of *M. athalia* from the Balkan, that was found to be genetically distinct from all other studied specimens, including some other specimens from Balkan (Figure 3b). The clusters for the rest of the *M. athalia* and *M. celadussa* were found to overlap to some extent in PCA plots for both datasets (Figure 3b). The analysis of genomic admixture using

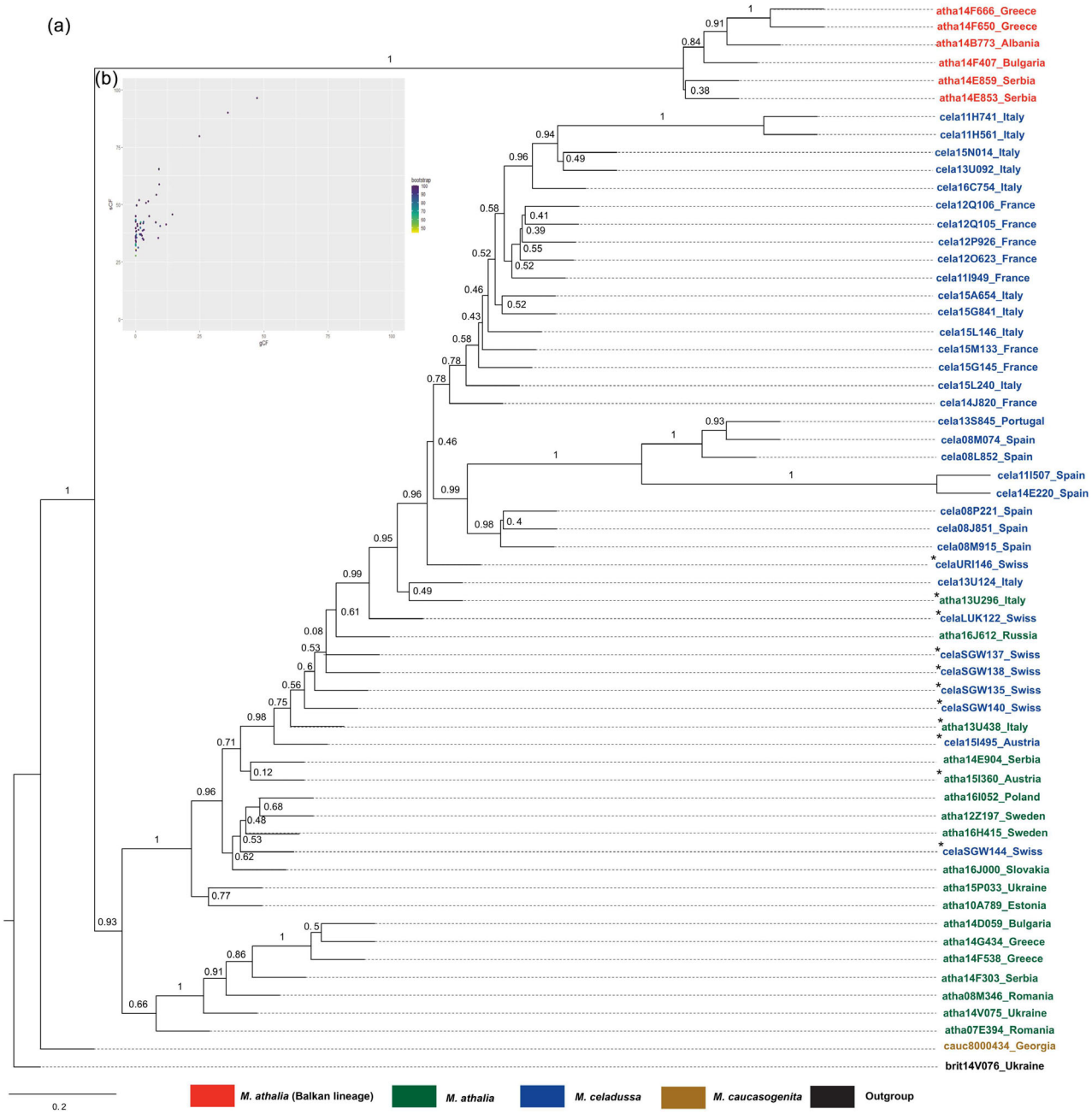

FIGURE 2 (a) ASTRAL species tree based on the target enrichment TE60 dataset and rooted with *M. britomartis* where numbers on the branches are quartet support scores. Contact zone specimens are indicated with the asterisks (b) sCF versus gCF plot for the TE60 dataset

STRUCTURE produced the highest likelihood estimate for $K = 2$ clusters for the ddRAD dataset, consistent with the results obtained by Tahami et al. (2021) and $K = 3$ for the target enrichment dataset (Figure S4). To better visualize the patterns of admixture spatially, we also mapped the membership coefficient matrix obtained from Bayesian clustering at $K = 3$ for target enrichment to the geographic coordinates plotted on a map in R v 4.0.4 (R core team, 2021). The comparison of STRUCTURE barplots from $K = 2$ to $K = 4$ for both datasets (Figure 4a) shows the distinctiveness of the Balkan *M. athalia* clade. These contrast with the presence of admixture in many of the *M. athalia* specimens (both in the contact zone and notably far from it) and in several *M. celadussa* individuals from the contact zone. This pattern is also evident from the geographic distribution of genomic admixture (Figure 4b), where the Balkan clade, some *M. athalia* (in the Balkans, Eastern Europe and N. Scandinavia) and the *M. celadussa* from south-western Europe (Spain, France and most of Italy) are shown to have pure gene pools (green, orange and pink coloured pies on the map, respectively), whereas the individuals from the contact zone and from a wide area in central Europe showed considerable genomic admixture (Figure 4b).

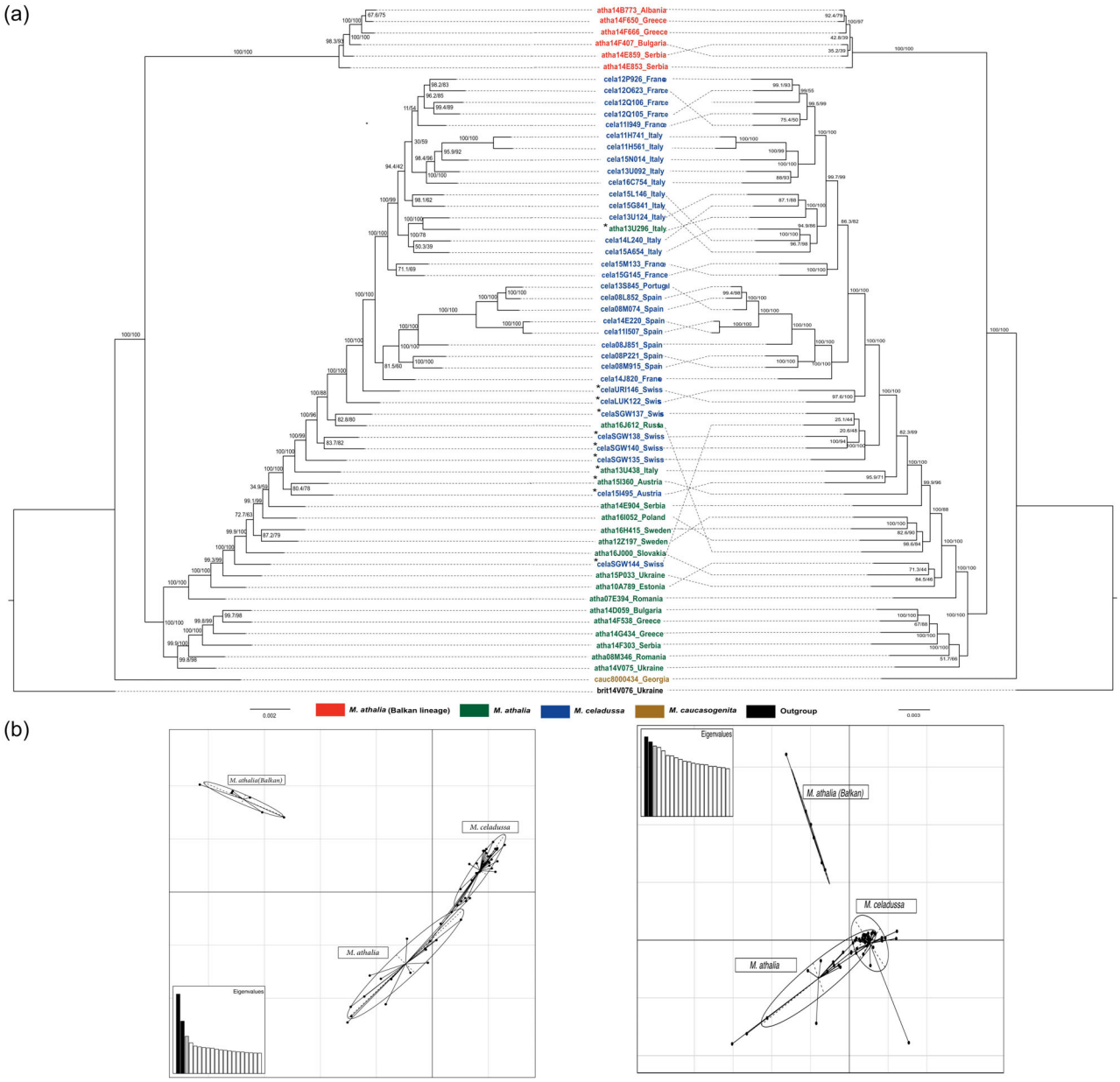

FIGURE 3 (a) Comparison of maximum-likelihood phylogenetic trees obtained from the TE60 dataset (on the left) and ddRAD dataset (on the right). Both trees are rooted with *Melitaea britomartis*. Numbers on the branches indicate SH-aLRT/UFBoot support values calculated based on 1000 replicates. Contact zone specimens are indicated with the asterisks (b) PCA plots for target enrichment SNP dataset (on left, PC1 and PC2 with variances 6.93% and 4.57%, respectively) and ddRAD dataset (on right, PC1 and PC2 with variances 3.46% and 3.23%, respectively).

Comparison - species delimitation

Based on MLE (Table 3), the three-species scenario achieved the highest rank among the two scenarios tested for both datasets (Table 3). However, as per BF scale, the BF values for both two-species (Run D, lumping *M. athalia* except Balkan lineage and *M. celadussa* together) and three-species scenarios (Run C, Balkan *M. athalia*, *M. athalia* and *M. celadussa*) were greater than 10, hence were concluded as decisive for both target enrichment and ddRAD datasets (Table 3). From the tr2 analysis, negative log likelihood values −3,659,835.67, −2,101,731.61 and −123,421.81 were calculated for the null model, model1 (with two species, *M. athalia* except Balkan lineage and *M. celadussa* as one species and Balkan lineage of *M. athalia* as another) and model2 (with three species, *M. athalia* excluding Balkan lineage, *M. celadussa* and Balkan lineage of *M. athalia* as separate species) respectively. Thus, the likelihood is distinctly higher for the three species model.

DISCUSSION

Using a parapatric pair of *Melitaea* butterflies as a model system, we compared two genomic approaches for their utility in elucidating patterns of genetic structure and admixture, and investigated whether

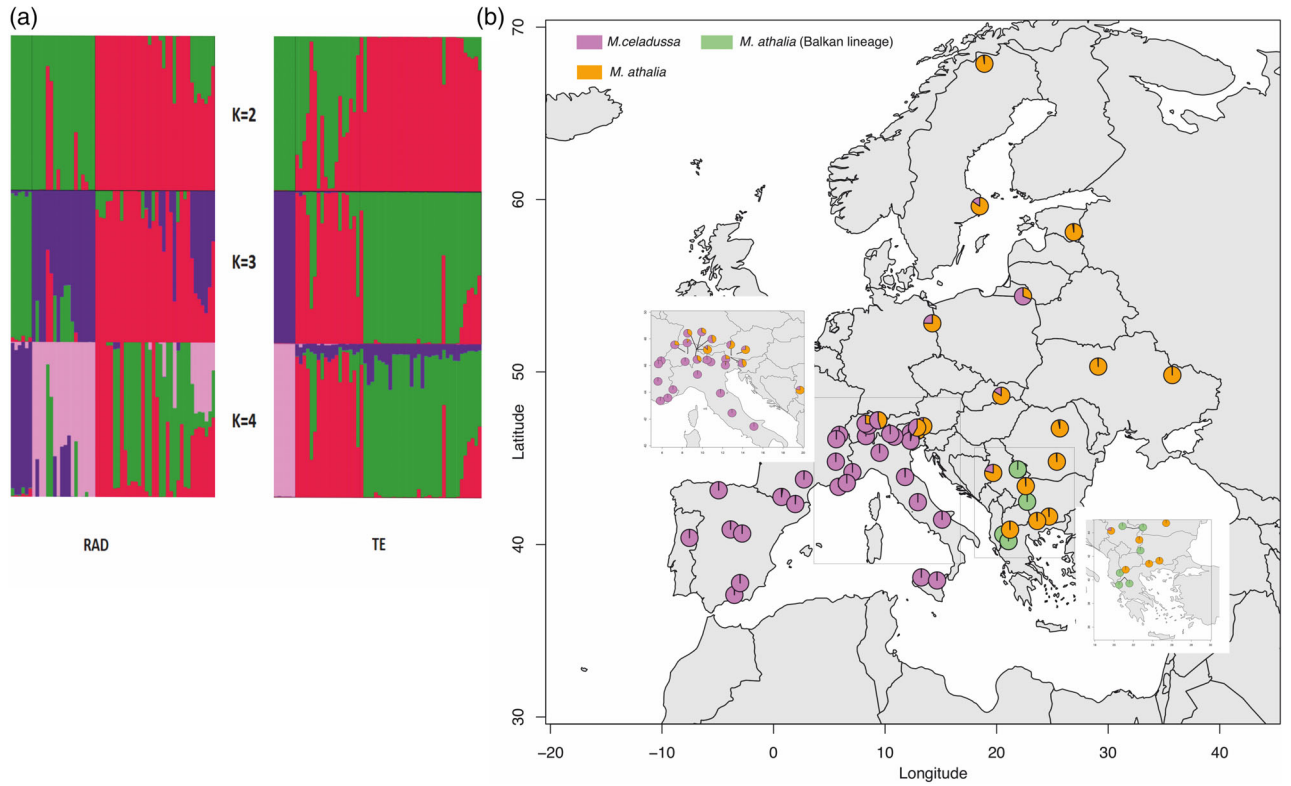

FIGURE 4 (a) STRUCTURE analysis of *Melitaea athalia-Melitaea celadussa* for $\Delta K = 5$ for the target enrichment dataset (right) and ddRAD dataset (left). The aligned barplots show cluster assignments at 2–4, from top to bottom. (b) Pie charts based on membership coefficient matrix (q-matrix) at $K = 3$ for the target enrichment dataset mapped on geographic coordinates.

TABLE 3 Comparison of different species delimitation models. Calculation of Bayes factor followed this tutorial—BFD-tutorial-1.pdf (netdna-cdn.com)

	Scenarios	Description	Number of species	MLE	Bayes factor	Rank
target enrichment	Current taxonomy (Run B)	((A, BA), C)	2	−44,476.8	NA	3
	Three species (Run C)	(A, C, BA)	3	−44,286.4	380.74	1
	Two species (Run D)	((A, C), BA)	2	−44,301.2	351.3	2
ddRAD	Current taxonomy (Run B)	((A, BA), C)	2	−175,641.7	NA	3
	Three species (Run C)	(A, C, BA)	3	−169,271.3	12,740.8	1
	Two species (Run D)	((A, C), BA)	2	−170,593.7	10,096	2

Note: BF = 2*(MLE_1 – MLE_0), where MLE_1 represents and alternate scenario and MLE_0 represents current taxonomy.
Abbreviations: A, *M. athalia* excluding Balkan lineage; BA = Balkan lineage of *M. athalia*; C, *M. celadussa*.

genome-wide datasets could provide means for well-informed and robust species delimitation under parapatry. Both sequence capture and RADseq have been independently explored for species delimitation before (Gueuning et al., 2020; Pante et al., 2015; Smith et al., 2014), but to the best of our knowledge, our study is among the first to focus on species delimitation in a parapatric system (Linck et al., 2019 being a rare exception). Additionally, only few studies have compared two genomic approaches in the same study system (Harvey et al., 2016; Leaché et al., 2015; Manthey et al., 2016), all of them utilizing UCEs as a primary sequence capture method. Consequently, there has been little discussion over the benefits and drawbacks of various genomics methods in species delimitation.

In our study system, both the ML inference and ASTRAL species tree inference recovered the same set of major clades in both the target enrichment and ddRAD approaches, thus providing a consistent phylogenetic picture for the group. *M. celadussa* was long considered as a subspecies of *M. athalia* (Higgins, 1955) but was recently given a full species status (Leneveu et al., 2009; Wiemers et al., 2018). We recovered *M. celadussa* as a non-monophyletic group and most of *M. athalia* as a paraphyletic grade with respect to *M. celadussa*. Specimens for which intermediate genital characters were reported (and which according to DNA barcodes are *M. celadussa*) were also part of this paraphyletic *M. athalia-M. celadussa* grade. A higher discordance among gene trees was observed for the *M. athalia-M. celadussa* grade

as per gCF analyses, the conflict possibly arising due to introgression of *M. celadussa* genes into *M. athalia*. Also, most differences between ddRAD and target enrichment trees were found for this grade as per *tipdiff* plot. Comparison of patterns of admixture using STRUCTURE analyses revealed different optimum *K* values for the two methods. However, the PCA plots looked largely similar, with a distinct Balkan lineage cluster and two overlapping clusters of *M. athalia* and *M. celadussa*. The barplot for STRUCTURE analyses at $K = 2$ for ddRAD and $K = 3$ for target enrichment, as well as the geographic distribution of genomic admixture (for target enrichment $K = 3$), show patterns consistent with those observed by Tahami et al. (2021), with Balkan *M. athalia* and south-western European *M. celadussa* having non-admixed gene pools and a high admixture in individuals from the contact zone and in *M. athalia* in C. Europe. One notable difference is that, in the case of the target enrichment dataset, the clusters tended to be purer. For example, the Balkan clade and *M. celadussa* specimens from south-western Europe are clearly shown as non-admixed in the target enrichment-based barplot and admixture pie plots mapped on geographic coordinates. This could be attributed to less noise in the target enrichment dataset, owing to low amount of missing data and off-target loci (Table 2). For the species delimitation analyses, both the three-species and two-species scenarios received a decisive support by target enrichment and ddRAD, suggesting that both datasets supported each of the two scenarios equally. However, testing of alternate species assignments using tr2 delimitation for target enrichment TE60 data favoured the three-species hypothesis over the two-species one. Thus, no single hypothesis was consistently supported by all of the different methods.

Based on ddRAD data, Tahami et al. (2021) hypothesized that the taxa examined here could have originated from three main refugia in Europe along several glacial cycles: Iberian and Italian Peninsula for *M. celadussa*; the Balkan Peninsula for the Balkan lineage; and an eastern refugium for non-admixed *M. athalia*. They also hypothesized that the observed patterns could have resulted from adaptive introgression of a set of *M. celadussa* genes eastward into *M. athalia*. Based on the results from our current study, we see that two main genetically distinct entities are identified (the Balkan lineage of *M. athalia* and the rest of the *M. athalia* and *M. celadussa* specimens) which agrees with the two-species scenario. But *M. athalia* and *M. celadussa* are mainly identified as separate species by their large distribution, genital differentiation and mitochondrial and nuclear genetic separation. In addition to that, there is an unexpected split between two lineages of *M. athalia* where the genetically diverging Balkan lineage stays distinct from the rest of *M. athalia* from this region despite geographical proximity, which could be a possible explanation for the three-species scenario.

Target enrichment versus ddRAD – benefits and pitfalls

In target enrichment methods the sequencing efforts are concentrated on a pre-defined set of loci, and therefore high coverage and little missing data is expected (Mayer et al., 2021). Different target enrichment methods mainly differ in the gene regions they target and, as researchers can select loci at desired levels of divergence (Banker et al., 2020), they have been shown to be generally efficient in elucidating affinities at both deep and shallow phylogenetic scales (Bagley et al., 2020; Espeland et al., 2018; Hamilton et al., 2016).

Using a recently developed target enrichment kit that targets about 2900 CDS regions (Mayer et al., 2021), we were able to recover many gene loci across the individuals even after stringent filtering criteria (1002 gene loci for TE60) which shows the capture experiment to be successful. We found that the number of loci captured was sufficient to uncover the phylogenetic relationships of *Melitaea* irrespective of whether loci were concatenated and assumed to have the same genealogical history, or if they were allowed to have independent histories and analysed within a summary coalescent framework using ASTRAL. This demonstrates that the target enrichment loci contain sufficient phylogenetic information that could be useful to delimit closely related taxa. Also, the amount of missing data in both target enrichment datasets was very low compared to the ddRAD dataset.

RADseq methods are inexpensive and quick compared to target enrichment and do not need genomic resources to design probes. However, they are most useful at shallow scales of divergence, as the number of homologous loci obtained decreases with phylogenetic distance (Lee et al., 2018; Wagner et al., 2013). Furthermore, the locus dropout effect also increases with divergence, leading to non-random patterns of missing data (Arnold et al., 2013). RADseq methods also require high quality material, that is, better preserved tissue to successfully extract sufficient amounts of DNA, while it is possible to recover genomic data from old museum material using target enrichment (Call et al., 2021; Mayer et al., 2021).

It has been suggested that the method of data collection should not influence the results of phylogenetic analyses at intermediate levels of divergence (Manthey et al., 2016). From our comparison of target enrichment and ddRAD, this should also be the case at an alpha taxonomic level. Therefore, the choice of method to tackle species delimitation at such levels could be entirely up to the researcher based on resources available and quality of the samples. Here, some practical aspects are worth taking into consideration: target enrichment is generally more expensive than RADseq because of the costs associated with the library preparation and purchasing enrichment probes (Harvey et al., 2016). The target enrichment laboratory work tends to be slower due to additional hybridization and enrichment steps although with the newer kits it is now significantly faster. However, as the loci obtained from RADseq are random and not known, different RAD datasets are hard to combine and are not comparable due to their unique nature. Different RAD datasets are also likely to be composed of loci with different levels of conservativeness (Lee et al., 2018). Target enrichment has an important advantage in terms of cross-compatibility, since the targeted loci remain fixed. As the loci being recovered are already known in target enrichment and usually have low levels of missing data, this could be an advantage in identifying a universal set of markers useful for species delimitation. It would

also enable a comparison across different species groups, as well as combining different datasets, and hence provide means for standardized delimitation of, for example, allopatric populations and other settings where delimitation of species is inherently difficult.

Recently, the use of a standard set of nuclear markers for species delimitation and identification was proposed (Eberle et al., 2020), which can be studied in and compared across all animals and would allow disentangling recently diverged lineages. This idea was tested by Dietz et al. (2021) using empirical data from several metazoan lineages. They found that so-called Universal Single-Copy Orthologs (USCOs) performed better than DNA barcodes to delimit closely related species, irrespective of the assembly approach or tree reconstruction method used (Dietz et al., 2021). A similar idea of having a unified set of loci for phylogenomic and population genetic studies has also been explored by Singhal et al. (2017), but exclusively for squamates. Basing species delimitation on a standard set of molecular markers would be an important step toward a more stabilized taxonomy, particularly under conditions where delimitation is bound to be arbitrary. This concerns the delimitation of allopatric and parapatric populations, as well as asexual strains, but would likely turn out to provide efficient means for a delimitation in all settings.

The conundrum of parapatric species delimitation

Finding universally applicable criteria for species delimitation is rendered extremely challenging particularly due to the complexity of biological systems (Eberle et al., 2020) and the semantics around the meaning of the term 'species' (de Queiroz, 2007). We find reaching a consensus over the definition of species not foreseeable and hence do not discuss this further here, but we are slightly more optimistic that the disagreement over the epistemological aspects of species delimitation could be overcome by adopting efficient and standardizable approaches such as those presented by Eberle et al. (2020) and Dietz et al. (2021).

Parapatric taxa constitute an unusually complex case for species delimitation, because they typically show frequent hybridization in the contact zone (Bull, 1991; Hewitt, 1988) and genetic admixture that may extend far beyond the zone of contact (Johnson et al., 2015; Osada et al., 2010). Parapatric taxa tend to be morphologically and ecologically very similar, and morphologically intermediate individuals may occur (Guiller et al., 2017; Saino & Villa, 1992; Slender et al., 2017). From a temporal perspective, parapatric systems are presumably usually young—although initial differentiation may be much older (Ebdon et al., 2021)—and may have undergone either slow merging of populations or increased differentiation, the latter being promoted by disruptive selection and reinforcement (Barton & Hewitt, 1985). These complexities also characterize our study system, which we assume is resulting from a postglacial secondary contact of populations differentiated in two refugia during the last glaciation, and possibly in earlier ones (Tahami et al., 2021).

As the delineation of parapatric taxa is inherently arbitrary and largely subjective, delimiting admixing parapatric taxa in a consistent manner would promote taxonomic stability. Optimally, delimitation should be based on broad genomic data, because it is little affected by choice of marker and provides sufficient data to obtain a proper understanding of evolutionary relationships and admixture between the populations. Generating a species list at global, continental or local scale for any group of organisms would require a widely accepted pragmatic solution for identifying parapatric taxa either as infraspecific or as full species. Parapatric systems are also characterized by notable discordance between different types of characters across individuals. Such discordance, both mitonuclear and mito-morphological, was also observed in *Melitaea*. Keeping in mind that admixing taxa or hybrids cannot be assigned to species by definition and that around the hybrid zone most specimens might fall into this category, we would need to identify which parameters are most relevant to the formation of evolutionary distinct units (frequency of hybrids, mitonuclear discordance, presence and frequency of morphological intermediates, distance from the contact zone where introgressed specimens can be found). A drawback of considering parapatric species as infraspecific is that this might prevent their recognition as biological entities worth of protection in countries where conservation legislation does not recognize the value of taxa below species level. However, the same holds true at all levels of biodiversity, including an uncountable number of genetically unique populations whose recognition as valuable units of biodiversity is seldom legally acknowledged.

CONCLUSIONS

The emergence of high-throughput genetic tools is revolutionizing systematic research at all phylogenetic levels. Many platforms and protocols to generate genomic-scale datasets are available, each with their specific strengths and shortcomings. This poses a question: which particular method provides the best means to address the research question at a given phylogenetic level? In this work, we focused on answering this question at the interface of population and phylogenetic levels, using a particularly challenging system of two parapatric butterfly species with frequent introgression and widespread admixture as a model. While both ddRAD sequencing and target enrichment methods provided largely congruent pictures of the evolutionary history of our study system, the latter has the benefit of providing higher levels of scalability. Compared to ddRAD datasets, sequence capture is characterized by a dramatically lower degree of missing data due to a smaller locus dropout effect and sequencing of a predefined set of loci at a high coverage. Presently, there is no consensus over the principles for delimitation of parapatric taxa. We argue that this is inherently based on arbitrary criteria and reaching a widely accepted consensus is desperately needed. As patterns of admixture like those observed here are more a rule than an exception in parapatric systems, we propose to regard the admixing parapatric taxa in a consistent way. Considering them regularly either as subspecies or full species would promote taxonomic stability, but both solutions would also be characterized by some shortcomings.

AUTHOR CONTRIBUTIONS

Mukta Joshi: Data curation (equal); formal analysis (lead); investigation (lead); validation (lead); visualization (lead); writing – original draft (lead). **Marianne Espeland:** Data curation (equal); methodology (lead); project administration (equal); supervision (equal); writing – review and editing (equal). **Vlad Dincă:** Conceptualization (equal); resources (equal); writing – review and editing (equal). **Roger Vila:** Conceptualization (equal); resources (equal); writing – review and editing (equal). **Mohadeseh Sadat Tahami:** Formal analysis (supporting); investigation (supporting). **Lars Dietz:** Methodology (equal); writing – review and editing (equal). **Christoph Mayer:** Methodology (equal); writing – review and editing (equal). **Sebastian Martin:** Methodology (equal). **Leonardo Dapporto:** Formal analysis (supporting); resources (equal); writing – review and editing (equal). **Marko Mutanen:** Conceptualization (equal); funding acquisition (lead); project administration (equal); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

This work was supported by Academy of Finland (grant no. 314702) to Marko Mutanen. Roger Vila was supported by Grant PID2019-107078GB-I00 funded by MCIN/AEI/ 10.13039/501100011033. Vlad Dincă was supported by the Academy of Finland (Academy Research Fellow, decision no. 328895). The authors would like to thank Sandra Kukowka for her help in the lab work. We would like to acknowledge CSC–IT Center for Science, Finland, for providing computational resources. We would also like to thank Tomochika Fujisawa for helpful discussion on tr2 analysis. We would like to thank three anonymous reviewers, whose comments greatly improved the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw target enrichment sequencing data is archived on the NCBI Sequence Read Archive under BioProject ID PRJNA802033. Raw ddRAD data are available under BioProject ID PRJNA638526.

ORCID

Mukta Joshi https://orcid.org/0000-0002-6785-8306
Marianne Espeland https://orcid.org/0000-0002-6800-4783
Vlad Dincă https://orcid.org/0000-0003-1791-2148
Roger Vila https://orcid.org/0000-0002-2447-4388
Lars Dietz https://orcid.org/0000-0001-6469-381X
Christoph Mayer https://orcid.org/0000-0001-5104-6621
Sebastian Martin https://orcid.org/0000-0003-3171-4420

REFERENCES

Arnold, B., Corbett-Detig, R.B., Hartl, D. & Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22, 3179–3190. https://doi.org/10.1111/mec.12276

Aronesty, E. (2011) *ea-utils*: "*Command-line tools for processing biological sequencing data*". Available from: https://github.com/ExpressionAnalysis/ea-utils. Accessed 19th December 2021.

Bagley, J.C., Uribe-Convers, S., Carlsen, M.M. & Muchhala, N. (2020) Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Molecular Phylogenetics and Evolution*, 152, 106769. https://doi.org/10.1016/j.ympev.2020.106769

Banker, S.E., Lemmon, A.R., Hassinger, A.B., Dye, M., Holland, S.D., Kortyna, M.L. et al. (2020) Hierarchical hybrid enrichment: multitiered genomic data collection across evolutionary scales, with application to chorus frogs (*Pseudacris*). *Systematic Biology*, 69, 756–773. https://doi.org/10.1093/sysbio/syz07

Barton, N.H. & Hewitt, G.M. (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16, 113–148.

Beuret, H. (1933) Zur Kenntnis der Variabilität von Melitaea athilia Rott., und psuedathalia Rev. (Lepidoptera). *Journal of the Swiss Entomological Society*, 15, 78–94. https://doi.org/10.5169/seals-400770

Breinholt, J.W., Earl, C., Lemmon, A.R., Lemmon, E.M., Xiao, L. & Kawahara, A.Y. (2018) Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored Phylogenomics. *Systematic Biology*, 67, 78–93. https://doi.org/10.1093/sysbio/syx048

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

Bull, C.M. (1991) Ecology of parapatric distributions. *Annual Review of Ecology and Systematics*, 22, 19–36.

Call, E., Mayer, C., Twort, V., Dietz, L., Wahlberg, N. & Espeland, M. (2021) Museomics: Phylogenomics of the moth family Epicopeiidae (Lepidoptera) using target enrichment. *Insect Systematics and Diversity*, 5, 6. https://doi.org/10.1093/isd/ixaa021

Chernomor, O., von Haeseler, A. & Minh, B.Q. (2016) Terrace aware data structure for Phylogenomic inference from Supermatrices. *Systematic Biology*, 65, 997–1008. https://doi.org/10.1093/sysbio/syw037

Dapporto, L., Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M. et al. (2019) Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Molecular Ecology Resources*, 19, 1623–1636. https://doi.org/10.1111/1755-0998.13059

Dietz, L., Dömel, J.S., Leese, F., Mahon, A.R. & Mayer, C. (2019) Phylogenomics of the longitarsal Colossendeidae: The evolutionary history of an Antarctic Sea spider radiation. *Molecular Phylogenetics and Evolution*, 136, 206–214. https://doi.org/10.1016/j.ympev.2019.04.017

Dietz, L., Eberle, J., Mayer, C., Kukowka, S., Bohacz, C., Baur, H., et al. (2021) Standardized nuclear markers advance metazoan taxonomy. *BioRxiv*, 2021.05.07.443120. https://doi.org/10.1101/2021.05.07.443120

Earl, D.A. & von Holdt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Ebdon, S., Laetsch, D.R., Dapporto, L., Hayward, A., Ritchie, M.G., Dincă, V. et al. (2021) The Pleistocene species pump past its prime: Evidence from European butterfly sister species. *Molecular Ecology*, 30, 3575–3589. https://doi.org/10.1111/mec.15981

Eberle, J., Ahrens, D., Mayer, C., Niehuis, O. & Misof, B. (2020) A plea for standardized nuclear markers in metazoan DNA taxonomy. *Trends in Ecology & Evolution*, 35, 336–345. https://doi.org/10.1016/j.tree.2019.12.003

Espeland, M., Breinholt, J., Willmott, K.R., Warren, A.D., Vila, R., Toussaint, E.F.A. et al. (2018) A comprehensive and dated phylogenomic analysis of butterflies. *Current Biology*, 28, 770–778. https://doi.org/10.1016/j.cub.2018.01.061

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

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, 717–726. https://doi.org/10.1093/sysbio/sys004

Fujisawa, T., Aswad, A. & Barraclough, T.G. (2016) A rapid and scalable method for multilocus species delimitation using Bayesian model comparison and rooted triplets. *Systematic Biology*, 65, 759–771. https://doi.org/10.1093/sysbio/syw028

Gavrilets, S., Li, H. & Vose, M.D. (2000) Patterns of parapatric speciation. *Evolution*, 54, 1126–1134. https://doi.org/10.1111/j.0014-3820.2000.tb00548.x

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W. et al. (2009) Solution hybrid selection with ultralong oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, 27, 182–189. https://doi.org/10.1038/nbt.1523

Gueuning, M., Frey, J.E. & Praz, C. (2020) Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in central European bees. *Molecular Ecology*, 29, 4203–4220. https://doi.org/10.1111/mec.15629

Guiller, G., Lourdais, O. & Ursenbacher, S. (2017) Hybridization between a euro-Siberian (*Vipera berus*) and a Para-Mediterranean viper (*V. aspis*) at their contact zone in western France. *Journal of Zoology*, 302, 138–147. https://doi.org/10.1111/jzo.12431

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307–321. https://doi.org/10.1093/sysbio/syq010

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M. & Bond, J.E. (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, 16, 1–20. https://doi.org/10.1186/s12862-016-0769-y

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*, 65, 910–924. https://doi.org/10.1093/sysbio/syw036

Hendry, A.P., Bolnick, D.I., Berner, D. & Peichel, C.L. (2009) Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, 75, 2000–2036. https://doi.org/10.1111/j.1095-8649.2009.02419.x

Herrera, S. & Shank, T.M. (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *Molecular Phylogenetics and Evolution*, 100, 70–79. https://doi.org/10.1016/j.ympev.2016.03.010

Hewitt, G.M. (1988) Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*, 3, 158–167. https://doi.org/10.1016/0169-5347(88)90033-X

Higgins, L.G. (1955) Descriptive catalogue of the genus Mellicta bill-berg (Lepidoptera: Nymphalidae) and its species, with supplementary notes on the genera Melitaea and Euphydryas. *Transactions of the Royal Entomological Society of London*, 106, 1–127. https://doi.org/10.1111/j.1365-2311.1888.tb00660.x

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. https://doi.org/10.1093/molbev/msx281

Hohenegger, J. (2014) Species as the basic units in evolution and biodiversity: Recognition of species in the recent and geological past as exemplified by larger foraminifera. *Gondwana Research*, 25, 707–728. https://doi.org/10.1016/j.gr.2013.09.009

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jarman, S.N. & Elliott, N.G. (2000) DNA evidence for morphological and cryptic cenozoic speciations in the Anaspididae, "living fossils" from the triassic. *Journal of Evolutionary Biology*, 13, 624–633. https://doi.org/10.1046/j.1420-9101.2000.00207.x

Johnson, B.B., White, T.A., Phillips, C.A. & Zamudio, K.R. (2015) Asymmetric introgression in a spotted salamander hybrid zone. *Journal of Heredity*, 106, 608–617. https://doi.org/10.1093/jhered/esv042

Jombart, T. & Ahmed, I. (2011) Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Kendall, M., Almagro-Garcia, J. & Colijn, C. (2017) Treespace: statistical exploration of landscapes of phylogenetic trees. *Molecular Ecology Resources*, 17, 1385–1392. https://doi.org/10.1111/1755-0998.12676

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589. https://doi.org/10.1038/nmeth.4285

Kass, R. & Raftery, A. (1995) Bayes factors. *Journal of American Statistical Association*, 90, 773–795.

Knowlton, N. (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics*, 24, 189–216. https://doi.org/10.1146/annurev.ecolsys.24.1.189

Kück, P. & Longo, G.C. (2014) FASconCAT-G: Extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, 11, 1–8. https://doi.org/10.1186/s12983-014-0081-x

Lanfear, R. (2018) *Calculating and interprating gene- and site-concordance factors in phylogenomics*. Available from: http://www.robertlanfear.com/blog/files/concordance_factors.html. Accessed on 19th December 2021.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, 14, 1–14. https://doi.org/10.1186/1471-2148-14-82

Leaché, A. D. & Bouckaert, R. R. (2018) *Species trees and species delimitation with SNAPP: A tutorial and worked example*. Available from: http://evomicsorg.wpengine.netdna-cdn.com/wp-content/uploads/2018/01/BFD-tutorial-1.pdf. Accessed on 19th December 2021.

Leaché, A.D., Fujita, M.K., Minin, V.N. & Bouckaert, R.R. (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, 63, 534–542. https://doi.org/10.1093/sysbio/syu018

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D. & Linkem, C.W. (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, 7, 706–719. https://doi.org/10.1093/gbe/evv026

Lee, K.M., Kivelä, S.M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N. et al. (2018) Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multilocus sanger data in a species-rich moth genus. *Systematic Biology*, 67, 925–939. https://doi.org/10.1093/sysbio/syy029

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, 61, 727–744. https://doi.org/10.1093/sysbio/sys049

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*, 97, 346–361. https://doi.org/10.1111/j.1095-8312.2009.01208.x

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools.

Bioinformatics, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linck, E., Epperly, K., van Els, P., Spellman, G.M., Bryson, R.W., McCormack, J.E. et al. (2019) Dense geographic and genomic sampling reveals Paraphyly and a cryptic lineage in a classic sibling species complex. *Systematic Biology*, 68, 956–966. https://doi.org/10.1093/sysbio/syz027

Lischer, H.E.L. & Excoffier, L. (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299. https://doi.org/10.1093/bioinformatics/btr642

Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20, 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Manthey, J.D., Campillo, L.C., Burns, K.J. & Moyle, R.G. (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: A test in cardinalid tanagers (Aves, genus: *Piranga*). *Systematic Biology*, 65, 640–650. https://doi.org/10.1093/sysbio/syw005

Mayer, C., Sann, M., Donath, A., Meixner, M., Podsiadlowski, L., Peters, R. S. et al. (2016) BaitFisher: A software package for multispecies target DNA enrichment probe design. *Molecular Biology and Evolution*, 33, 1875–1886. https://doi.org/10.1093/molbev/msw056

Mayer, C., Dietz, L., Call, E., Kukowka, S., Martin, S. & Espeland, M. (2021) Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens. *Systematic Entomology*, 46, 649–671. https://doi.org/10.1111/syen.12481

Minh, B.Q., Hahn, M.W. & Lanfear, R. (2020a) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37, 2727–2733. https://doi.org/10.1093/molbev/msaa106

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

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B. & Worm, B. (2011) How many species are there on earth and in the ocean? *PLoS Biology*, 9, 1–8. https://doi.org/10.1371/journal.pbio.1001127

Mutanen, M., Hausmann, A., Hebert, P.D.N., Landry, J.F., de Waard, J.R. & Huemer, P. (2012) Allopatry as a Gordian knot for taxonomists: patterns of DNA barcode divergence in Arctic-alpine Lepidoptera. *PLoS One*, 7, e47214. https://doi.org/10.1371/journal.pone.0047214

Osada, N., Uno, Y., Mineta, K., Kameoka, Y., Takahashi, I. & Terao, K. (2010) Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. *Molecular Ecology*, 19, 2884–2895. https://doi.org/10.1111/j.1365-294X.2010.04687.x

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C. et al. (2015) Use of RAD sequencing for delimiting species. *Heredity*, 114, 450–459. https://doi.org/10.1038/hdy.2014.105

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7, e37135. https://doi.org/10.1371/journal.pone.0037135

Platania, L., Menchetti, M., Dincă, V., Corbella, C., Kay-Lavelle, I., Vila, R. et al. (2020) Assigning occurrence data to cryptic taxa improves climatic niche assessments: biodecrypt, a new tool tested on European butterflies. *Global Ecology and Biogeography*, 29, 1852–1865. https://doi.org/10.1111/geb.13154

Poelstra, J.W., Ellegren, H. & Wolf, J.B.W. (2013) An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Nature*, 111, 467–473. https://doi.org/10.1038/hdy.2013.68

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

de Queiroz, K. (2005) Ernst Mayr and the modern concept of species. *Proceedings of National Academy of Sciences*, 102, 6600–6607. https://doi.org/10.1073/pnas.0502030102

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

Rosenberg, N.A. (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology*, 4, 137–138. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Saino, N. & Villa, S. (1992) Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. *The Auk*, 109, 543–555.

Saino, N., Lorenzini, R., Fuscot, G.E.A., Randii, E., Biologia, D. & Bo, O.E. (1992) Genetic variability in a hybrid zone between carrion and hooded crows (*Corvus corone corone* and *C. c. Cornix*, Passeriformes, Aves) in North-Western Italy. *Biochemical Systematics and Ecology*, 20, 605–613.

Singhal, S., Grundler, M., Colli, G. & Rabosky, D.L. (2017) Squamate conserved loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources*, 17, e12–e24. https://doi.org/10.1111/1755-0998.12681

Slender, A.L., Louter, M., Gardner, M.G. & Kleindorfer, S. (2017) Patterns of morphological and mitochondrial diversity in parapatric subspecies of the thick-billed Grasswren (*Amytornis modestus*). *Emu - Austral Ornithology*, 117, 264–275. https://doi.org/10.1080/01584197.2017.1313686

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C. & Brumfield, R.T. (2014) Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, 63, 83–95. https://doi.org/10.1093/sysbio/syt061

Tahami, M.S., Dincă, V., Lee, K.M., Vila, R., Joshi, M., Heikkilä, M. et al. (2021) Genomics reveal admixture and unexpected patterns of diversity in a parapatric pair of butterflies. *Genes*, 12, 2009. https://doi.org/10.3390/genes12122009

Tewksbury, J.J., Anderson, J.G.T., Bakker, J.D., Billo, T.J., Dunwiddie, P.W., Groom, M.J. et al. (2014) Natural history's place in science and society. *Bioscience*, 64, 300–310. https://doi.org/10.1093/biosci/biu032

Van Oorschot, H., Coutsis, J. G. (2014). The genus *Melitaea* Fabricius, 1807 (Lepidoptera: Nymphalidae, Nymphalinae). Taxonomy and systematics with special reference to the male genitalia. Pardubice: Tshikolovets Publications.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L. et al. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, 787–798. https://doi.org/10.1111/mec.12023

Wiemers, M., Balletto, E., Dinca, V., Fric, Z.F., Lamas, G., Lukhtanov, V. et al. (2018) An updated checklist of the European butterflies (Lepidoptera, Papilionoidea). *ZooKeys*, 811, 9–45. https://doi.org/10.3897/zookeys.811.28712

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 15–30. https://doi.org/10.1186/s12859-018-2129-y

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Table S1 A list of specimens included in the SNAPP analysis

Figure S1: Species tree generated using the concatenation approach in IQTREE with bootstrap/gCF/sCF values labelled on branches

Figure S2: Dotplot of tip differences between TE60 and ddRAD trees obtained using the R function *tipdiff*.

Table S2: Specimens and their membership by location

Figure S3: PCA plot for unlinked SNPs from ddRAD dataset

Figure S4: The ΔK plot for the ddRAD dataset showing optimum at $K = 2$ (on right) and at $K = 3$ (on left) for the target enrichment dataset.

Figure S5: ASTRAL species tree based on the target enrichment TE30 dataset and rooted with *M. britomartis* where numbers on the branches are quartet support scores. Contact zone specimens are indicated with the asterisks.

How to cite this article: Joshi, M., Espeland, M., Dincă, V., Vila, R., Tahami, M.S., Dietz, L. et al. (2022) Delimiting continuity: Comparison of target enrichment and double digest restriction-site associated DNA sequencing for delineating admixing parapatric *Melitaea* butterflies. *Systematic Entomology*, 47(4), 637–654. Available from: https://doi.org/10.1111/syen.12557