



HAL
open science

Patterns of speciation in a parapatric pair of *Saturnia* moths as revealed by target capture

Maria Khan, Mukta Joshi, Marianne Espeland, Peter Huemer, Carlos Lopez-Vaamonde, Marko Mutanen

► **To cite this version:**

Maria Khan, Mukta Joshi, Marianne Espeland, Peter Huemer, Carlos Lopez-Vaamonde, et al.. Patterns of speciation in a parapatric pair of *Saturnia* moths as revealed by target capture. *Molecular Ecology*, 2023, 10.1111/mec.17194 . hal-04273563v2

HAL Id: hal-04273563

<https://hal.inrae.fr/hal-04273563v2>

Submitted on 7 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Patterns of speciation in a parapatric pair of *Saturnia* moths as revealed by target capture

Maria Khan¹  | Mukta Joshi¹  | Marianne Espeland²  | Peter Huemer³  |
Carlos Lopez-Vaamonde^{4,5}  | Marko Mutanen¹ 

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

²Leibniz Institute for the Analysis of
Biodiversity Change, Bonn, Germany

³Tiroler Landesmuseen
Betriebsges.m.b.H.,
Naturwissenschaftliche Sammlungen, Hall,
Austria

⁴INRAE, URZF, Orléans, France

⁵IRBI, UMR 7261, CNRS-Université de
Tours, Tours, France

Correspondence

Mukta Joshi, Ecology and Genetics
Research Unit, University of Oulu, PO
Box 3000, FI-90014 Oulu, Finland.
Email: mukta.joshi@oulu.fi

Funding information

Academy of Finland, Grant/Award
Number: 314702; COFUND, Grant/
Award Number: 101081280; Suomen
Kulttuurirahasto, Grant/Award
Number: 00230503; European Regional
Development Fund (ERDF), Grant/Award
Number: projectInfoBioS and EX011185

Handling Editor: J. A. H. Benzie

Abstract

The focus of this study has been to understand the evolutionary relationships and taxonomy of a widely distributed parapatric species pair of wild silk moths in Europe: *Saturnia pavonia* and *Saturnia pavoniella* (Lepidoptera: Saturniidae). To address species delimitation in these parapatric taxa, target enrichment and mtDNA sequencing was employed alongside phylogenetic, admixture, introgression, and species delimitation analyses. The dataset included individuals from both species close to and farther away from the contact zone as well as two hybrids generated in the lab. Nuclear markers strongly supported both *S. pavonia* and *S. pavoniella* as two distinct species, with hybrids forming a sister group to *S. pavoniella*. However, the Maximum Likelihood (ML) tree generated from mtDNA sequencing data presented a different picture, showing both taxa to be phylogenetically intermixed. This inconsistency is likely attributable to mitonuclear discordance, which can arise from biological factors (e.g., introgressive hybridization and/or incomplete lineage sorting). Our analyses indicate that past introgressions have taken place, but that there is no evidence to suggest an ongoing admixture between the two species, demonstrating that the taxa have reached full postzygotic reproductive isolation and hence represent two distinct biological species. Finally, we discuss our results from an evolutionary point of view taking into consideration the past climatic oscillations that have likely shaped the present dynamics between the two species. Overall, our study demonstrates the effectiveness of the target enrichment approach in resolving shallow phylogenetic relationships under complex evolutionary circumstances and that this approach is useful in establishing robust and well-informed taxonomic delimitations involving parapatric taxa.

KEYWORDS

genomics, parapatry, speciation, species delimitation, target capture

Maria Khan and Mukta Joshi contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

A fundamental unit of biological diversity is the species. Essentially, all organisms are assignable to a species, although the identification of a species is not always straightforward due to morphological or genetic similarities between species. These difficulties are further confounded by the fact that many species remain undescribed and hence unnamed. The semantics over the word 'species' represents another long-standing debate among taxonomists. Consequently, over 20 species concepts have put forward (Hausdorf, 2011; Mallet, 2005; Mayden, 1997). These concepts, despite being mutually incompatible, share one common denominator: species are separately evolving metapopulation lineages (De Queiroz, 1999). Species delimitation, that is, the process of identifying species-level biological diversity (Carstens et al., 2013), is a crucial aspect of all taxonomic, evolutionary, and biodiversity research (Firneno et al., 2021; Hey, 2001).

Despite the importance of the species in characterizing biodiversity, delimiting species remains challenging as boundaries between them cannot always be defined unambiguously based on morphological and/or genetic characteristics (Barley et al., 2013). For example, many species tend to exhibit considerable phenotypic variability across their range, while others appear morphologically indistinguishable despite reproductive isolation (Knowlton, 1993). Evolutionary processes such as hybridization, introgression, and incomplete lineage sorting can further complicate the delimitation of species (Harrison & Larson, 2014; Ivanov et al., 2018), especially in cases of limited geographic sampling (Chambers & Hillis, 2020). Reaching a consensus over the principles on how species should be delimited under various evolutionary circumstances is a pre-requisite for efficient communication of biodiversity and conservation efforts (Gaston & Spicer, 2004; Tobias et al., 2010).

Climatic oscillations during the Pleistocene have shaped the genetic variability and distribution of European Lepidoptera to a considerable extent (Schmitt, 2007). As speciation is often a long and gradual process, the uncertainty over species boundaries is especially pronounced under allopatric and parapatric settings (Joshi et al., 2022; Mutanen et al., 2012; Nosil, 2008). In cases of parapatrically distributed taxa, delimitation of species is further complicated by the possibility of admixture and hybridization between taxa, as the ranges of two taxa may have a zone of overlap with an incomplete reproductive barrier (Bull, 1991). It is important to note that observing parapatry in nature is difficult and can easily but not necessarily be regarded as another mode of speciation since parapatric distribution can be the result of parapatric speciation or secondary contact between populations first diverged in allopatry (Abbott et al., 2013).

Parapatric and perhaps incompletely diverged taxa are valuable models to study the speciation process and character displacement through reinforcement (Pfennig, 2016). Some remarkable examples of parapatry involving European taxa include the fritillary butterflies *Melitaea athalia* Rottemburg, 1775 and *M. celadussa* Fruhstorfer, 1910 (Tahami et al., 2021), the skipper butterflies *Pyrgus malvae* Linnaeus, 1758 and *P. (malvae) melotis* Duponchel, 1832 (de Jong, 1987), the

Hooded Crow (*Corvus cornix*) and the Carrion Crow (*C. corone*) (Poelstra et al., 2013; Saino et al., 1992; Saino & Villa, 1992), as well as the fire-bellied toads *Bombina bombina* and *B. variegata* (Yanchukov et al., 2006). As new cases continue to be revealed, the secondary contacts of genetically differentiated populations are likely to be much more common in the nature than is presently appreciated.

Here, we focus on a parapatric pair of small Emperor moths belonging to the family Saturniidae: *Saturnia pavonia* Linnaeus, 1758 and *Saturnia pavoniella* Scopoli, 1763. The two species are widely distributed across Europe with *S. pavonia* found in Northern and Central Europe and *S. pavoniella* in Southern and Eastern Europe while the situations in Southern France, Iberian, and Balkan Peninsulas are still under debate (Huemer & Nässig, 2003; Mazel, 2007). *S. pavoniella* was separated from *S. pavonia* based on the differences in morphological characteristics of the wings and other body parts including the genitalia, as well as the infertility of female F1 hybrids. However, in Southern France, *S. pavoniella* populations show a mixture of morphological characters from both taxa (Mazel, 2007). Indeed, both species have been observed to overlap locally and create a suture zone in Southern France, Czech Republic, Italy, and Austria. While the distributions of the two taxa have not yet been elucidated in all detail, it seems evident that they form a pair of parapatrically distributed taxa with a long but narrow zone of contact, because both taxa are seldom co-occurring. The two taxa have been suggested to occasionally exchange genetic material as male hybrids can be fertile (Huemer & Nässig, 2003). Whether this capacity to hybridize has resulted in introgression and the disruption of the barrier to gene flow rendering the populations to remain discrete has remained unstudied in natural conditions.

With the increasing availability of genomic-scale data, it is now possible to understand the interplay between different evolutionary processes responsible for formation and distribution of biological diversity (Ferrer Obiol et al., 2023). Although whole genomes are now being sequenced at a relatively fast rate and are potentially an ideal source of data for phylogenomic studies, they remain unavailable for many non-model taxa (Breinholt et al., 2018). Approaches such as target enrichment methods allow for the efficient capture and sequencing of specific regions of the genome, which can be used to infer species boundaries (Mamanova et al., 2010). Further methodological advances have allowed the development of species delimitation programs in statistically rigorous frameworks such as Bayesian statistics and multispecies coalescent or MSC (Leaché et al., 2014; Yang, 2015; Zhang et al., 2018), to delimit species from molecular data.

Our aim was to shed light on the evolutionary relationships and speciation process of the parapatric and near-cryptic species pair of *S. pavonia* and *S. pavoniella* and to investigate whether they represent a single variable species or two distinct species in terms of biological and phylogenetic species concepts. To address these questions, we utilized a target enrichment approach that generates large number of fixed loci across the genome (Joshi et al., 2022; Mayer et al., 2021). In relation to that, we further investigated the levels of admixture and traces of historical or ongoing introgression between the two taxa. Motivation to understand this system stems largely from the

desire to find out better justified and more standardized ways to delimit species under complex evolutionary circumstances such as parapatry as this remains a largely unresolved problem that causes taxonomic instability. The benefits of using a set of standardized genetic markers in establishing a consistent criterion for species delimitation has been discussed in Eberle et al. (2020) and implemented in a set of metazoan taxa in Dietz et al. (2022).

2 | MATERIALS AND METHODS

2.1 | Taxon sampling and sample preparation

The study dataset involved a total of 27 specimens of the genus *Saturnia*, including 11 *S. pavonia*, 13 *S. pavoniella*, and 1 *S. josephinae* as an outgroup (Table 1). Additionally, two laboratory hybrid specimens (one male and one female) were also included in the dataset. One of the aims of including lab-reared hybrids was to confirm

them as hybrids based on genetic data. The specimens were initially identified and assigned to putative species based on morphological characteristics (Details given in Appendix S1). The samples were collected from nine different countries across Europe (Austria, Croatia, Czech Republic, Finland, France, Italy, Slovenia, Spain) using the sex pheromone (E6,Z11)-C16Ac synthesized for this study by Till Tolasch (Germany). Lures were prepared at INRAE Orléans from 11-mm red rubber septum lures (Wheaton Scientific) loaded with heptane solutions (100 µL, 1 mg/mL) of the synthesized pheromone that had been shipped by courier to INRAE from Tolasch Lab in Germany. Loaded septa were placed in a 20-ml glass vial and kept in a freezer when not in use. Lures were sent by post to several colleagues across Europe to attract and collect males for this study (see Acknowledgements for list of colleagues).

The female used to rear the hybrids was a lab reared *S. pavoniella* from the Var region in Southern France and the male was a wild *S. pavonia* from the Loire valley. The crossing was made in a garden in Indre & Loire in central France where the reared *S. pavoniella*

TABLE 1 Specimen metadata and dataset statistics.

Sample ID	Species	Country of collection	SRA accession	No. of loci recovered	% Missing data
SAT003	<i>Saturnia pavonia</i>	France	SRR23477969	453	25.40
SAT004	<i>Saturnia pavonia</i>	France	SRR23477968	422	29.44
SAT006	<i>Saturnia pavoniella</i>	France	SRR23477957	409	31.58
SAT008	<i>Saturnia pavonia</i>	France	SRR23477949	427	28.82
SAT009	<i>Saturnia pavoniella</i>	France	SRR23477948	456	24.32
SAT010	<i>Saturnia pavoniella</i>	France	SRR23477947	425	29.00
SAT011	<i>Saturnia pavoniella</i>	France	SRR23477946	441	27.27
SAT012	<i>Saturnia pavoniella</i>	France	SRR23477945	456	24.00
SAT013	<i>Saturnia pavonia</i>	France	SRR23477944	404	30.47
SAT020	<i>Saturnia pavonia</i>	Spain	SRR23477943	458	23.63
SAT021	<i>Saturnia pavonia</i>	Spain	SRR23477967	430	26.46
SAT022	Hybrid <i>pavonia</i> X <i>pavoniella</i>	France	SRR23477966	474	21.15
SAT023	Hybrid <i>pavonia</i> X <i>pavoniella</i>	France	SRR23477965	471	21.23
SAT024	<i>Saturnia pavonia</i>	Spain	SRR23477964	383	35.08
SAT026	<i>Saturnia pavonia</i>	Czechia	SRR23477963	479	20.78
SAT028	<i>Saturnia pavoniella</i>	Czechia	SRR23477962	459	23.27
SAT031	<i>Saturnia pavoniella</i>	Slovenia	SRR23477961	475	20.96
TLMF Lep 03073	<i>Saturnia pavoniella</i>	Austria	SRR23477960	390	33.83
TLMF Lep 30352	<i>Saturnia pavonia</i>	Austria	SRR23477959	422	28.69
TLMF Lep 30357	<i>Saturnia pavoniella</i>	Austria	SRR23477958	407	30.73
TLMF Lep 30358	<i>Saturnia pavoniella</i>	Austria	SRR23477956	431	27.03
TLMF Lep 30361	<i>Saturnia pavoniella</i>	Italy	SRR23477955	407	30.06
TLMF Lep 30362	<i>Saturnia pavoniella</i>	Italy	SRR23477954	425	27.95
MM27425	<i>Saturnia pavonia</i>	Finland	SRR23477953	447	24.91
MM27428	<i>Saturnia pavonia</i>	Finland	SRR23477952	433	27.93
SAT033	<i>Saturnia pavoniella</i>	Croatia	SRR23477950	483	18.56
SAT034	<i>Saturnia josephinae</i>	Spain	SRR23477951	496	20.82

female was tied around the base of its two forewings with a piece of 50–70 cm of cotton string. The string was itself attached to a shrub to prevent the female from escaping. The female called and mated with one of the wild *S. pavonia* males she had attracted. The mated *S. pavoniella* female laid eggs and the hybrid caterpillars were reared to obtain F1 hybrid adults.

DNA was extracted from the antenna or thorax of either ethanol preserved or dry specimens. DNA extraction was performed using the QIAGEN DNeasy Blood and Tissue Kit (Hildesheim, Germany) or the Omega Bio-tek E.Z.N.A. Insect DNA Kit (Norcross, United States) following the protocols given by the manufacturers.

2.2 | Target enrichment library preparation

Following extraction, the DNA concentration of each sample was measured using the Invitrogen Quant-iT PicoGreen dsDNA (Waltham, United States) quantification kit. Approximately 100 ng of absolute DNA was taken for further processing as per the standard Agilent protocol; for the samples with low DNA yield, approximately 50 ng of absolute DNA was used, depending on the available quantity. For target enrichment protocol, the genomic DNA (gDNA) was fragmented using the standard protocols for mechanical DNA shearing on a Diagenode Bioruptor®, to achieve the average fragment size between 150 and 200 bp. The fragmented DNA was further taken for library preparation using SureSelect XT HS2 library prep kit which included end repair reaction and ligation of adenine residue to the 3' end of the blunt fragments (A-tailing) to allow ligation of barcoded adaptors. PCR amplification of adaptor-ligated libraries was then performed, after indexing with the Agilent SureSelect XT HS2 primer pairs; we used eight PCR cycles for all the samples, irrespective of the amount of absolute DNA taken initially. After the library construction, custom Agilent SureSelect baits (6–11.9 Mb) were used for solution-based target enrichment of a pool containing 8 or 16 libraries. We used LepZFMK 1.0 kit (Mayer et al., 2021) which targets 2953 CDS regions in 1753 nuclear genes, to enrich the gene regions of interest. The hybridization was performed on each pool with the Agilent SureSelect XT HS2 Target Enrichment Kit ILM module following the manufacturer's instructions. Enriched libraries were then captured with MyOne Streptavidin T1 beads. These captured libraries were PCR amplified and the final concentration of each captured library was measured using the Invitrogen Quant-iT PicoGreen dsDNA quantification kit. Following the enrichment, pooled libraries were sequenced using the Illumina Nextseq 500 platform (mid output) to generate paired-end 150-bp reads at Biomedicum Functional Genomics Unit (FuGU), Helsinki Institute of Life Science at University of Helsinki.

2.3 | Barcoding

For the mitochondrial DNA (mtDNA) sequencing, the *COI* gene was amplified using the primers HybLCO and HybHCO. PCR was

conducted following the protocols in deWaard et al. (2008) (protocol given in Appendix S1). The PCR cleanup was then performed using Sephadex columns (Sigma-Aldrich). The purified product was sent to MacroGen Europe for sequencing.

The *COI* sequences received from MacroGen Europe were aligned using ClustalW as implemented in MEGA X software v10.0.5 (Kumar et al., 2018) and BioEdit 7.2.5 sequence alignment editor (Hall, 1999). The reference sequence against which the other sequences were aligned was retrieved from BOLD Systems 4.0 database (<http://www.boldsystems.org>). This reference specimen was one of those included in the sample specimens. For the specimens which failed the sequencing, we extracted the *COI* barcode region from the target enrichment data using *exonerate* ver. 2.4.0 (Slater & Birney, 2005), with the alignment of individuals that were successfully barcoded as a query and our de-novo assembled target enrichment assemblies as a target (details of the command are given in Appendix S1). These extracted barcodes were aligned with the rest of the barcodes using Muscle aligner on EMBL-EBI server (<https://www.ebi.ac.uk/Tools/msa/muscle/>) with default parameters. The alignment was manually checked in Geneious Prime (<https://www.geneious.com/prime/>) before proceeding for phylogenetic analysis. All the newly generated barcodes were uploaded to BOLD database (DOI: <https://doi.org/10.5883/DS-PAVONI>).

To visualize divergence patterns in *COI*, phylogenetic trees of the aligned *COI* barcode sequences were generated using IQTREE ver. 2.0.3 (Minh et al., 2020), using the best-fit model estimated by the program using ModelFinder (Kalyanamoorthy et al., 2017) and ultrafast bootstrapping with 1000 replicates. We used the *-bnni* option to reduce the impact of model violation and overestimated branch support. Twenty likelihood searches were performed and the tree with the highest log-likelihood was visualized in FigTree v1.4.4 (available from tree.bio.ed.ac.uk/software/figtree/).

2.4 | Target enrichment bioinformatics

The demultiplexed data after being checked for quality were processed using the TENriAn pipeline (Mayer et al., 2021). During the first step of the TENriAn workflow, adaptor sequences and low-quality bases were removed from the demultiplexed data using *fastp* (Chen et al., 2018). The reads were then assembled using Trinity version 2.9.0 (Grabherr et al., 2011). We skipped the Contamcheck step during the assembly, where the data are checked and filtered for potential cross-contaminations, using 'all against all' BLAST searches as it was observed to be removing large amounts of useful data due to our samples being very closely related. Instead, we used the original assembled files before contamination checking as input for Orthograph v. 0.7.1. (Petersen et al., 2017) step. During this step, clusters of orthologous loci are inferred, using a pre-generated reference database. In our case, this reference database was generated using the coding gene alignments of five species – *Bombyx mori*, *Danaus plexippus*, *Heliconius melpomene*, *Melitaea cinxia*, and *Papilio glaucus*. Next, the orthologous clusters were aligned, using

reference Hidden Markov Models (HMMs) in hmalign (part of the HMMER package, <http://hmmerr.org>), and the alignments were filtered using various criteria such as coverage and outliers, following Mayer et al. (2021). We further removed the loci having GC content of higher than 60%, as those can have a negative effect on downstream phylogenetic analysis.

2.5 | SNP calling

To perform population genetic analyses, we also did variant calling on the target enrichment dataset. We used the ingroup sample with the highest number of orthologous clusters inferred overall (during orthology assessment step in the TEnriAn workflow) as a reference for SNP calling. The raw cleaned data (i.e., after removing adaptors and low-quality bases) were mapped against this reference using the BWA-MEM algorithm in bwa 0.7.17 (available from bio-bwa.sourceforge.net) with the minimum seed length set to 30.

SNPs were then extracted from the sorted BAM files generated during mapping by using the samtools v. 1.16.1 (Danecek et al., 2021) *mpileup* option piped together with the bcftools *call* option. Filtering was then accomplished using bcftools to remove indels and multiallelic SNPs and keeping only biallelic SNPs. The number of SNPs obtained after this filtering step were 10,265. The SNP coverage distribution is shown in Figure S6.

2.6 | Phylogenetic analyses

The filtered alignments from the TEnriAn workflow were concatenated using FASConcat-G available at <https://github.com/PatrickKueck/FASconCAT-G>. This concatenated dataset was used to infer Maximum Likelihood tree using IQTREE 2.0.3 (Minh et al., 2020). To partition the data according to loci, we set up partitioned analysis using IQTREE with the option -m TESTMERGEONLY to resemble PartitionFinder (Chernomor et al., 2016) and the 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 for a partitioned analysis in IQ-TREE. We used the ultrafast bootstrap approximation with 1000 replicates (Hoang et al., 2018). To further reduce the impact of severe model violations, the -bnni option was used. We also performed an SH-like approximate likelihood ratio test (Guindon et al., 2010) with 1000 replicates using the -alrt option. Twenty tree searches were run and the tree with the highest log-likelihood was visualized in Figtree.

Generally, in phylogenomic datasets with multiple loci, some amount of discordance is expected among the gene trees. To investigate the impacts of incomplete lineage sorting, we inferred a species tree using ASTRAL-III v. 5.7.4 (Zhang et al., 2018), since it is statistically consistent under the multispecies coalescent model. A set of gene trees was inferred using IQTREE version 2.0.3 (Minh et al., 2020) where IQTREE was prompted to perform model selection (using ModelFinder) and tree inference separately for each

locus. To improve the ASTRAL results further, the gene trees were filtered using program TreeShrink (Mai & Mirarab, 2018) before running ASTRAL to detect and prune any abnormally long branches. The program was run in default 'per-species' mode with a false positive tolerance rate (α) set to 0.05. The resulting shrunk output gene trees were used as an input for ASTRAL, which generated a species tree along with the quartet score. The tree was visualized in FigTree v1.4.4.

2.7 | Population genetics

To get a notion of the possible number of genetic clusters and related species in the dataset, we performed Principal Component Analysis (PCA) on the target enrichment SNP dataset. The PCA was ran using *dudi.pca* function in the 'ade4' package (Jombart & Ahmed, 2011) in R studio v1.4.1106. We also calculated the pairwise F_{ST} from the same SNP dataset using R package *hierfstat* (Goudet, 2005).

STRUCTURE analysis (Pritchard et al., 2000) was also conducted on the target enrichment SNP dataset to infer population structure as well as admixture patterns. We used unlinked SNPs (as one of the assumptions of STRUCTURE includes that the loci are at HW equilibrium), and for this SNP data was filtered using a custom Perl script (<https://www.biostars.org/p/313701/>, accessed on 23/08/2023) to keep only one SNP per scaffold. The final number of SNPs after this filtering was 1831. The analysis was done assuming five clusters, that is for $K=1$ to $K=5$, each with 10 replicate runs. We used Clustering Markov Packager Across K or CLUMPAK ver. 1.1 (Kopelman et al., 2015) to align the cluster assignments across all 10 replicates for $K=2$, $K=3$, and $K=4$ and Distruct (Rosenberg, 2004) to show barplots for the K populations. Structure Harvester (Earl & vonHoldt, 2012) was used to obtain optimal number of genetic clusters (K) using ΔK method with 500,000 generations for the Markov chain and a value of 100,000 as burn-in. The program produces a graph based on the Evanno method (Evanno et al., 2005), between K cluster and DeltaK along the X and Y axis respectively, which shows the K with the highest DeltaK along the axes.

2.8 | Species delimitation analyses

To perform species delimitation analyses, we conducted Trinomial distribution analysis tr2 (Fujisawa et al., 2016), in which the distribution of rooted triplets is modelled to find the congruence between gene trees under multispecies coalescent framework. In short, it measures the incongruence of tree topologies within species and congruence between species. We tested a two-species hypothesis compared to the null model (assuming a single species) and obtained the results based on calculation of $-\log$ (posterior probability).

Additionally, we calculated the Genealogical Divergence Index (gdi) (Jackson et al., 2017) for the genetically distinct clades identified from phylogenetic analyses with BPP using the approach

described in Leaché et al. (2019). For this, we used our complete target enrichment dataset including all the loci. We first conducted a full BPP analysis (A11), where species tree and species delimitation are jointly estimated. We observed that BPP consistently inferred delimitation when all the populations were assigned to distinct species with posterior probability of 1, irrespective of prior combinations used. This asymptotic behaviour of Bayesian species delimitation models, that is, increase in the posterior probability of two-species model or species-splitting as the number of loci increases, has been discussed in Leaché et al. (2019). Next, we conducted a BPP analysis where species delimitation and species tree were fixed, to infer theta and tau values for each node corresponding to the distinct clades in the tree (BPP A00 analysis). We used beta prior values 0.004 for theta (population size) and 0.002 for tau (divergence times) respectively and the tree inferred by ASTRAL as a guide tree. Both values were estimated using Bayesian analysis, by specifying the 'e' option in the A00 analysis, which was conducted two independent times using a MCMC chain length of 1,000,000 and burn-in of 100,000. The convergence of the two runs was assessed in Tracer v1.7.2 (Rambaut et al., 2018) and the converged runs were concatenated to generate posterior distributions for the multispecies coalescent parameters. The gdi was calculated as

$$1 - e^{(-2\tau_{AB}/\theta_A)},$$

where tau_AB is the inferred divergence time of the clade from its sister group, and theta_A is the inferred population size for the clade. We determined the final theta and tau values as the median theta and tau values from the converged run. A gdi value of less than 0.2 suggests the conspecificity of sister clades, but values between 0.2 and 0.7 are indicative of the so-called 'grey zone', and a gdi value of >0.7 supports them to being a distinct species (Jackson et al., 2017).

2.9 | Isolation by distance

We tested for isolation by distance (IBD) using a Mantel test between a matrix of genetic distances and matrix of geographic distances using the *mantel.randtest* function in R package *adegenet*. This test finds the correlation between individual Edwards' genetic distances and Euclidean geographic distances (Mantel, 1967). Based on the simulated p-value, we determined whether isolation by distance was significant ($p < .05$).

2.10 | D-statistics

We used the program Dsuite (Malinsky et al., 2021) to calculate the Patterson's *D*-statistics to test for the presence of historical introgression in our dataset. This test was first introduced in Green et al. (2010) to detect the presence of Neandertal ancestry in modern humans, with further theoretical improvements

and applications in Reich et al. (2010) and Durand et al. (2011). This test computes the excess of alleles in one species (usually denoted as P3) in relation to two sister species (P1 and P2), and a fourth outgroup species as a reference. For this, we used a vcf file containing multiallelic SNPs as an input. The Dsuite program uses only biallelic SNPs by default, thereby eliminating the need to filter multiallelic SNPs file beforehand. Since we did not have three populations (excluding the outgroup), we ran this program assuming every specimen to be a distinct species/population, excluding hybrids from the analysis. *D*-statistics was calculated using *Dtrios* command, which calculates ABBA-BABA proportions and f4-ratio statistics for all possible trios of populations/species. We also performed the *f*-branch (fb) test which calculates *f*-branch statistic values for each branch on the input tree, including internal branches (Malinsky et al., 2021); it is useful in guiding the interpretation of correlated f4-ratio results. We used the tree inferred from ASTRAL as an input hypothesis tree, to calculate the *f*-branch statistic.

Additionally, we also tested for gene flow among the populations of two species using HyDe program (Blischak et al., 2018), which detects hybridization in phylogenomic datasets using phylogenetic invariants. The program uses *D*-statistics and site-pattern probabilities to test for hybridization. The analysis was run on concatenated sequence data. Since the main purpose behind using this program was to confirm the hybrid status of the two lab-reared hybrids based on genetic data, we defined the two hybrids as a separate population, along with *S. pavonia* and *S. pavoniella* as another two populations and *S. josephinae* as an outgroup. First, we performed a 'full analysis' using *run_hyde.py* script, which tests all possible triples of populations. Further, we tested for hybridization at individual level within populations detected as hybrids using *individual_hyde.py* script. The results were interpreted based on gamma (γ) value and a significance score (*p*-value).

3 | RESULTS

3.1 | Overview of the dataset

With regard to our genomic dataset, the average number of informative loci retained was 439 (SD=29.7, Table 1). The average amount of missing data was 26.42% (SD=0.043, Table 1). For our Mitochondrial DNA (mtDNA) dataset, 23 specimens were successfully sequenced, and for the remainder we used the barcodes extracted from target enrichment dataset for further analysis.

3.2 | Phylogenetic analyses

The preliminary specimen identification was done based on the morphological characteristics. The ML tree obtained from the mitochondrial COI dataset (Figure 2) did not show reciprocal monophyly, and the taxa seemed to be phylogenetically intermixed. However, five

specimens of *S. pavoniella* formed a sister group to the remaining specimens, suggesting that at least this species has species-specific haplotypes.

The ML tree from the target enrichment dataset showed two separate and strongly supported clusters corresponding to *S. pavonia* (indicated by orange, Figure 2) and *S. pavoniella* (indicated by blue, Figure 2), with SH-aLRT/UFBoot support values 100/100 and 100/99 respectively. The hybrids are highly supported as the sister group to *S. pavoniella*. One individual from *S. pavonia* (ID SAT024) has an unstable longer branch length which could be potentially caused by errors during the alignment step, or alternatively due to sequencing errors that were not filtered out at the bioinformatics step.

We observed that the species tree obtained from the dataset filtered using TreeShrink did not differ from the one obtained using data without filtering. Both species trees (generated from filtered and unfiltered gene trees) correspond to the ML tree generated from our genomic dataset (Figure 3 and Figure S5). Similar to the ML tree, *S. pavonia* and *S. pavoniella* were separated into two clusters with quartet support values of 0.99 and 1, respectively, with the hybrids highly supported as the sister group of *S. pavoniella*. As in the ML tree, this relationship is highly supported, but the clade including only the two hybrids has low support. We further observed that in both the COI and ASTRAL trees, two individuals from Italy, TLMF Lep 30361 and TLMF Lep 30362, were placed on an internal branch longer than for the rest of the individuals. To dissect the reasons behind this pattern, we further built a neighbour-joining (NJ) tree for the COI data, where it was observed that *S. josephinae* was nested within the variability of *S. pavonia* and *S. pavoniella*. Thus, we selected a more distant outgroup, *Saturnia pyri* Denis & Schiffermüller, 1775, to root the COI ML tree. However, we continued to observe the same pattern where the two specimens are placed on a longer internal branch, in ML tree. The neighbour-joining tree, on the other hand, once again displayed the pattern of *S. josephinae* being nested inside *S. pavonia*–*S. pavoniella* clade (Figure S7).

3.3 | PCA and F_{ST}

PCA based on the SNP dataset resulted in two genetically distinct clusters for *S. pavonia* and *S. pavoniella* (Figure 4). The two individuals that are situated between these two clusters in the figure are hybrids that appear intermediate and stay separated from both clusters (Figure 4). Three *S. pavonia* specimens with sample IDs. SAT026, MM27425, and MM27428 are separated from the rest of the *S. pavonia* individuals mainly along PC2, indicating to some extent the presence of genetic differentiation in these three individuals. The pairwise F_{ST} between the two species was calculated as 0.345.

3.4 | Genomic admixture and isolation by distance

The analysis of genomic admixture using STRUCTURE produced the highest likelihood estimate for $K=2$ clusters for the target

enrichment SNP dataset as can be seen in Table S1. The CLUMPAK analysis of STRUCTURE results gave bar-plots that show cluster assignment from $K=2$ to $K=4$ (Figure S1). The results do not suggest admixture between the two parapatric taxa. The only admixture that was observed was in the hybrids that were included in the study. From the geographic distribution of genomic admixture, it can be observed that both species have maintained genetic separation, despite range overlap in the hybrid zone (Figure 4). Only very slight admixture was observed (from $K=3$) in a few *S. pavonia* and *S. pavoniella* individuals (Figure S4).

The isolation by distance was observed to be non-significant (p -value $> .05$, Figure S3).

3.5 | Species delimitation

The tr2 species delimitation analysis gave log-likelihood scores based on posterior probability for the null model (assuming *S. pavonia* and *S. pavoniella* as a single species) and model1 (assuming *S. pavonia* and *S. pavoniella* as two different species) of $-(231271.46)$ and $-(7339.42)$ respectively, clearly preferring the two-species model (model1). The genealogical divergence index (gdi) value for divergence of *S. pavonia* clade from *S. pavoniella* was calculated as 0.277, suggesting the distinct species status of the two clades being ambiguous.

3.6 | Introgression

Out of 2024 trios analysed, high D -statistics values with significant p -values were observed for about 515 trios (D -statistics > 0.25 , Table S2). From these, we further separated the trios where both P1 and P2 are represented by same species, after which we obtained about 182 trios where both P1 and P2 are represented by *S. pavonia* and P3 by *S. pavoniella* and about 223 trios where P1 and P2 are represented by *S. pavoniella* and P3 by *S. pavonia*. Out of these, a significantly high D -statistic value (≥ 0.4) was observed for about 53 trios (Table 2) and additionally an elevated f_4 ratio (Figure S2). However, Malinsky et al. (2021) showed that the D and f_4 -ratio statistics are correlated and that a significantly elevated result for a trio does not necessarily indicate that the populations are involved in gene flow event. However, we observed stronger f -branch signals, that is, between 5 and 10% for several branches including pavonia7 (SAT026) and the entire pavoniella clade as well as pavoniella14 (SAT013) and several pavonia branches (Figure S2c), indicating past introgression incidents between the two species.

An analysis of hybridization using HyDe confirmed the hybrid status of the two lab-reared hybrid individuals, with gamma value of 0.43, which is the probability that the hybrid population is sister to population P2 (which was *S. pavoniella* in our case) and a p -value of less than 0.01. Further analysis at individual level revealed the significant presence of hybridization in one individual (SAT022), but surprisingly not in the other (SAT023, p -value $> .05$).

TABLE 2 D-statistics summary for SNP dataset.

P1	P2	P3	D-statistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
SAT020	SAT026	TLMF Lep 30362	0.49	10.19	2.30E-16	0.14	213.25	195.75	67
TLMF Lep 30361	SAT021	SAT020	0.49	8.69	2.30E-16	0.14	613.25	69.25	23.75
SAT020	SAT026	SAT031	0.49	9.39	2.30E-16	0.16	216.75	209.25	72.5
SAT020	SAT026	TLMF Lep 30357	0.49	9.19	2.30E-16	0.15	200.25	186	64.5
TLMF Lep 30361	SAT012	SAT021	0.48	6.36	1.97E-10	0.15	646	63.25	22
TLMF Lep 30362	SAT012	SAT021	0.48	6.41	142E-10	0.15	658.75	66.25	23.25
SAT020	SAT026	TLMF Lep 30361	0.48	9.16	2.30E-16	0.14	212.5	189.75	67.25
SAT020	SAT026	TLMF Lep 03073	0.46	8.60	2.30E-16	0.15	195.25	185.75	68.25
SAT020	SAT026	SAT033	0.46	9.24	2.30E-16	0.14	220.75	202.25	745
TLMF Lep 30361	SAT013	SAT020	0.46	7.47	7.75E-14	0.13	566.75	66.5	24.75
TLMF Lep 30362	SAT013	SAT020	0.45	7.11	1.13E-12	0.13	573.5	65.75	25
TLMF Lep 30362	SAT013	SAT021	0.44	5.53	3.29E-08	0.13	615.375	58.625	22.625
TLMF Lep 30362	SAT012	SAT020	0.44	6.83	8.51E-12	0.13	625	71	27.5
TLMF Lep 30361	SAT013	SAT021	0.44	5.25	1.53E-07	0.13	605.625	58.875	22.875
TLMF Lep 30361	SAT012	SAT008	0.44	5.25	1.51E-07	0.13	591.875	73.875	28.875
TLMF Lep 30361	SAT013	TLMF Lep 30,352	0.43	5.78	7.40E-09	0.12	528.125	69.875	27.625
SAT021	MM27425	TLMF Lep 03073	0.43	8.46	2.30E-16	0.12	176.875	163.125	64.625
SAT021	MM27425	TLMF Lep 30,362	0.43	8.80	2.30E-16	0.11	189	163.25	65
TLMF Lep 30361	SAT013	SAT008	0.43	6.54	6.27E-11	0.13	547.375	70.625	28.375
SAT021	MM27428	TLMF Lep 03073	0.43	5.74	9.71E-09	0.11	179.125	156.375	62.875
TLMF Lep 30362	SAT012	SAT008	0.43	5.08	3.82E-07	0.14	604.875	81.125	32.625
TLMF Lep 30362	SAT013	TLMF Lep 30,352	0.43	5.87	4.48E-09	0.12	536.875	68.625	27.625
SAT008	SAT026	SAT028	0.42	7.10	1.21E-12	0.15	218.875	197.375	80.125
TLMF Lep 30362	SAT028	SAT020	0.42	3.12	0.00182564	0.08	685.25	43.5	17.75
SAT020	SAT026	SAT009	0.42	9.40	2.30E-16	0.13	216.875	192.875	78.875
TLMF Lep 30361	SAT013	SAT003	0.42	5.31	1.10E-07	0.13	532.875	75.375	30.875
SAT033	SAT012	SAT021	0.42	4.81	1.48E-06	0.13	694.5	62.75	25.75
TLMF Lep 30362	TLMF Lep 30358	SAT021	0.42	3.41	0.00065793	0.10	703.25	47.5	19.5
SAT033	SAT012	SAT020	0.42	6.35	2.09E-10	0.13	663	71.25	29.25
TLMF Lep 03073	SAT013	SAT021	0.42	4.59	4.36E-06	0.13	603.125	58.625	24.125
SAT021	MM27425	SAT028	0.42	10.70	2.30E-16	0.12	179.625	176.125	72.625
TLMF Lep 03073	SAT013	SAT003	0.42	5.25	1.48E-07	0.13	524.125	74.875	30.875

TABLE 2 (Continued)

P1	P2	P3	D-statistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
SAT033	SAT012	SAT008	0.41	6.08	1.19E-09	0.13	640.625	80.625	33.375
SAT021	MM27428	TLMF Lep 30,362	0.41	5.25	1.48E-07	0.10	195.25	153.75	64
TLMF Lep 30361	TLMF Lep 30358	SAT020	0.41	4.99	5.97E-07	0.08	659.875	46.375	19.375
SAT020	SAT026	TLMF Lep 30,358	0.41	9.68	2.30E-16	0.13	195.125	186.625	78.125
TLMF Lep 30361	SAT012	SAT004	0.41	5.21	1.91E-07	0.12	566.875	70.125	29.375
SAT021	MM27425	TLMF Lep 30361	0.41	7.08	1.45E-12	0.10	185.875	159.125	66.875
TLMF Lep 30361	SAT011	SAT004	0.41	4.99	5.97E-07	0.12	530.125	67.375	28.375
SAT008	SAT026	SAT031	0.41	7.05	1.76E-12	0.13	237.125	194.375	81.875
SAT021	MM27428	SAT028	0.41	5.84	5.07E-09	0.12	186.125	164.375	69.375
SAT021	SAT004	TLMF Lep 03073	0.41	5.94	2.79E-09	0.10	204.375	137.625	58.125
TLMF Lep 30361	SAT013	SAT004	0.41	5.54	3.04E-08	0.12	528.625	65.375	27.625
SAT021	SAT004	SAT028	0.40	7.21	5.77E-13	0.10	200.125	147.375	62.625
SAT004	SAT026	SAT031	0.40	6.40	1.58E-10	0.12	226.125	174.875	74.375
TLMF Lep 30362	SAT013	SAT008	0.40	6.13	8.66E-10	0.12	554.125	66.625	28.375
SAT021	MM27428	TLMF Lep 30357	0.40	5.36	8.43E-08	0.11	175.5	158.75	67.75
TLMF Lep 30357	SAT012	SAT021	0.40	5.34	9.42E-08	0.13	644.125	61.125	26.125
SAT021	MM27425	SAT031	0.40	8.65	2.30E-16	0.11	192	17425	74.5
SAT031	SAT012	SAT021	0.40	4.80	1.57E-06	0.13	677.75	67.75	29
SAT021	SAT003	TLMF Lep 30362	0.40	7.46	8.63E-14	0.09	224.75	143.5	61.5
TLMF Lep 30361	SAT012	SAT003	0.40	4.08	4.42E-05	0.11	576.25	70	30

4 | DISCUSSION

S. pavonia and *S. pavoniella* represent a well-known system of recently diverged Lepidoptera with parapatric distributions and a narrow contact zone in Europe, similar to *M. athalia* and *M. celadussa* (Joshi et al., 2022; Tahami et al., 2021). We used this species pair as a model to elucidate the evolutionary genetic patterns and admixtures as well as species delimitation under parapatric mode of distribution. Based on hundreds of loci, we attempt to interpret the observations from evolutionary and taxonomic points of view, also taking into consideration the past climate oscillations that have shaped the genetic diversity of many European Lepidoptera (Hewitt, 1996, 2000, 2004). Systems like the one studied by us are likely to be common in nature, and while some cases are well documented, many more remain poorly understood and even undetected, because parapatric taxa usually represent recent divergence of lineages at the interface of species and population levels, rendering observations on two taxonomically distinct but possibly cryptic entities challenging. Possibly numerous closely related species have undergone a similar phase of a secondary contact following partial or full reproductive incompatibility that evolved during prolonged isolation, especially during glaciation maxima when ranges were split into several isolated refugia (Schmitt, 2007). Periodic expansions of the species' ranges due to climate oscillations can be expected to have led to many secondary contacts between diverged populations, which in turn would have provided grounds for ecological differentiation of reproductively diverged populations through reinforcement and character displacement.

4.1 | Mitonuclear discordance

The Maximum Likelihood tree for the mtDNA dataset did not fully separate the two taxa, suggesting that *S. pavonia* and *S. pavoniella* may not be distinct at the species level (Figure 1). In contrast, the

Maximum Likelihood tree and the ASTRAL species tree from the genomic dataset recovered the two taxa as separate entities, with hybrids being grouped together and appearing intermediate with regards to the wild collected specimens (Figures 1 and 2). The causes for the (seemingly) mitonuclear discordance can be non-biological (operational) or biological (Ivanov et al., 2018). Possible biological causes include incomplete lineage sorting (ILS), introgression, horizontal gene transfer, and *Wolbachia*-mediated selective sweeps following hybridization (Bergthorsson et al., 2003; Mutanen et al., 2016; Soucy et al., 2015; Toews & Brelsford, 2012). The operational causes include misidentifications, contaminations, paralogues, NUMTs, and chimeric sequences, but also inconsistencies due to inaccurate taxonomy (Funk & Omland, 2003; Mutanen et al., 2016). In this study, operational causes were ruled out via a thorough analysis and investigation; furthermore, the results were re-analysed to systematically eliminate operational causes as contributing factors. Therefore, it seems evident that the two species display genuine mitonuclear discordance.

While it is difficult to distinguish between incomplete lineage sorting and introgression based on phylogenomic analyses (Funk & Omland, 2003), insights of whether ILS is present in the dataset or not can be obtained. Generally, when ILS is present, individual genealogies tend to be discordant with each other and the species tree, because under this scenario gene genealogies are supposed to vary in a stochastic manner (Degnan & Rosenberg, 2009; Maddison, 1997). We observed that both concatenation and coalescent approaches displayed similar relationships between the two taxa and the normalized quartet score from ASTRAL was calculated as 0.53. This score is an indicator of gene tree discordance due to ILS, although in some cases it also could be an indicative of gene tree error (Mirarab, 2019). We conclude that moderate ILS likely is present in our study species.

There are several well-documented cases of mitonuclear discordance resulting from historical introgression (Linnen & Farrell, 2007; Shaw, 2002; Sota & Vogler, 2001). STRUCTURE analysis and



FIGURE 1. 1 Representative images of live wild female and male (a) *S. pavoniella* and (b) *S. pavonia* specimens.

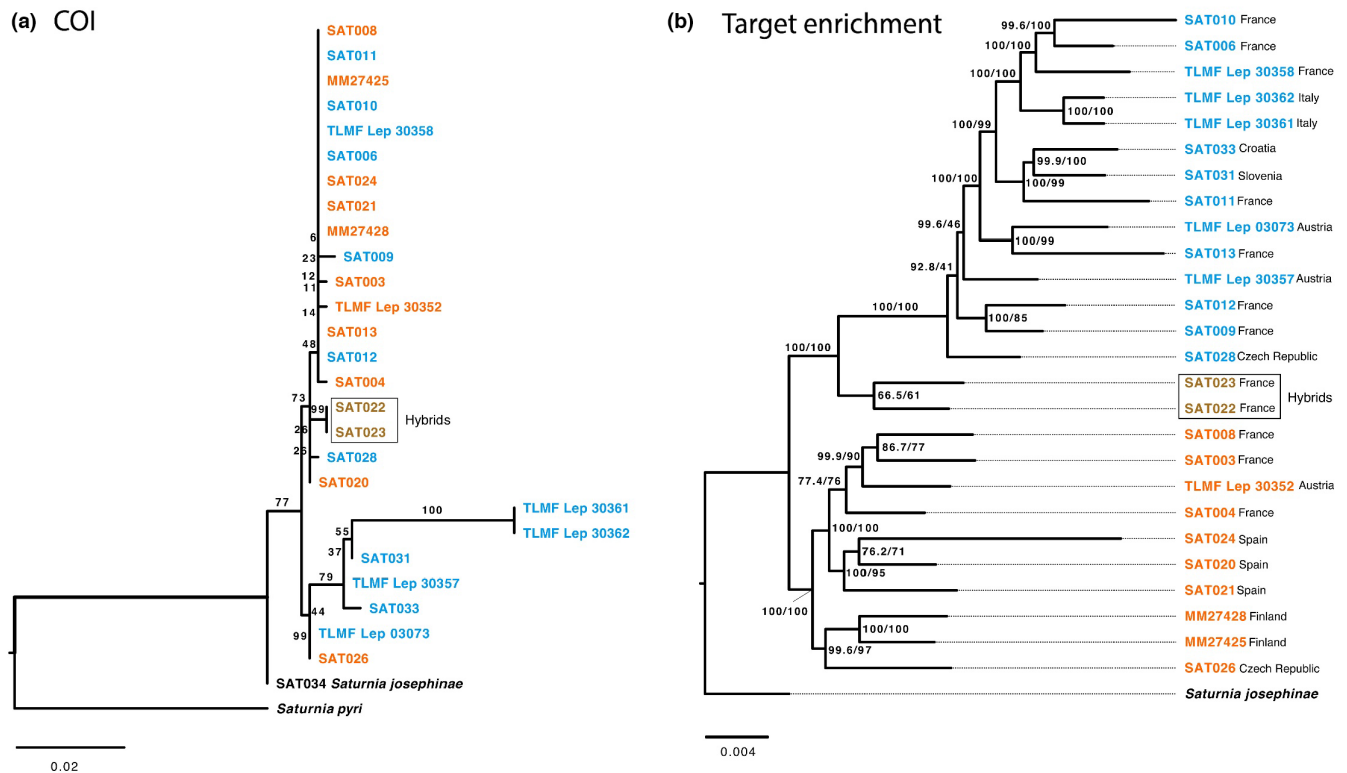


FIGURE 2 Maximum likelihood tree inferred using barcode data (on left) and target enrichment data (on right). *S. pavonia* individuals are indicated by orange colour and *S. pavoniella* by blue colour. For barcode ML tree, numbers on the branches indicate ultrafast bootstrap support values (UFBoot), and for target enrichment tree, they indicate sh-aLRT/UFboot values. Both specimens SAT022 (male) and SAT023 (female) indicated by the box are lab-reared F1 hybrids obtained from the following crossing: *pavoniella* female * *pavonia* male.

geographic distribution of genomic admixture demonstrate that *S. pavonia* and *S. pavoniella* do not display significant levels of admixture. It has been proposed that the two species could show sporadic introgression through fertile hybrid males (Huemer & Nässig, 2003). Analysis of ancient introgression using Dsuite supports this idea of episodes of introgression in the past. However, presently the two taxa show no or only little gene flow between them.

We show that the HyDe program can be reliably used to detect hybridization and the presence of hybrids in phylogenomic datasets. As in our case, the two hybrid individuals were lab-reared and their status as hybrids was therefore known. With wild collected specimens, hybrid status may be difficult to document, although its detection could sometimes be important for example in conservation efforts (Godinho et al., 2011; Harmoinen et al., 2021). As we demonstrated here, the gamma value can give an idea of F1 hybrids, as it represents the probability that the hybrid population is sister to one of the two parental populations. The value close to 0.5 suggests that they are about 50:50 (F1) hybrids (Blischak et al., 2018). At the same time, it should be emphasized that the process of hybridization is not always uniform, and that hence the gamma value can vary widely between individuals within hybridizing population. This was the case with our data.

Although the topologies of the barcode and ASTRAL trees differed from each other, in both trees we observed two specimens from Italy placed on long internal branches. In ASTRAL, such long

branches are generally indicative of gene tree discordance, as the branch lengths are in coalescent units (Sayyari & Mirarab, 2016). In some cases, they could also be indicative of the presence of pseudogenes or bad alignments due to missing data or shorter sequences, etc. As we had carefully checked all our alignments (barcode and nuclear genes) before proceeding for further analysis, the possibility of misalignments can be ruled out in both cases; the biological causes, if any, remain to be discovered. We further observed the somewhat surprising pattern of *COI* ML and NJ trees displaying different phylogenetic relationships when rooted on a more distant outgroup.

Results based on species delimitation analyses using BPP were found to contradict those of other analyses, in which *S. pavonia* and *S. pavoniella* were clearly identified as separate species. Gdi values have been shown to be unreliable predictors of the species status (Dietz et al., 2022), and gdi values estimated for taxonomically distinct species may sometimes be lower than those estimated for intraspecific entities (Jackson et al., 2017), which was also observed to be the case in the present study.

4.2 | Character displacement and parapatry

Results from our analysis of nuclear markers strongly suggest that the two taxa are separate species that fulfil the criteria of both the

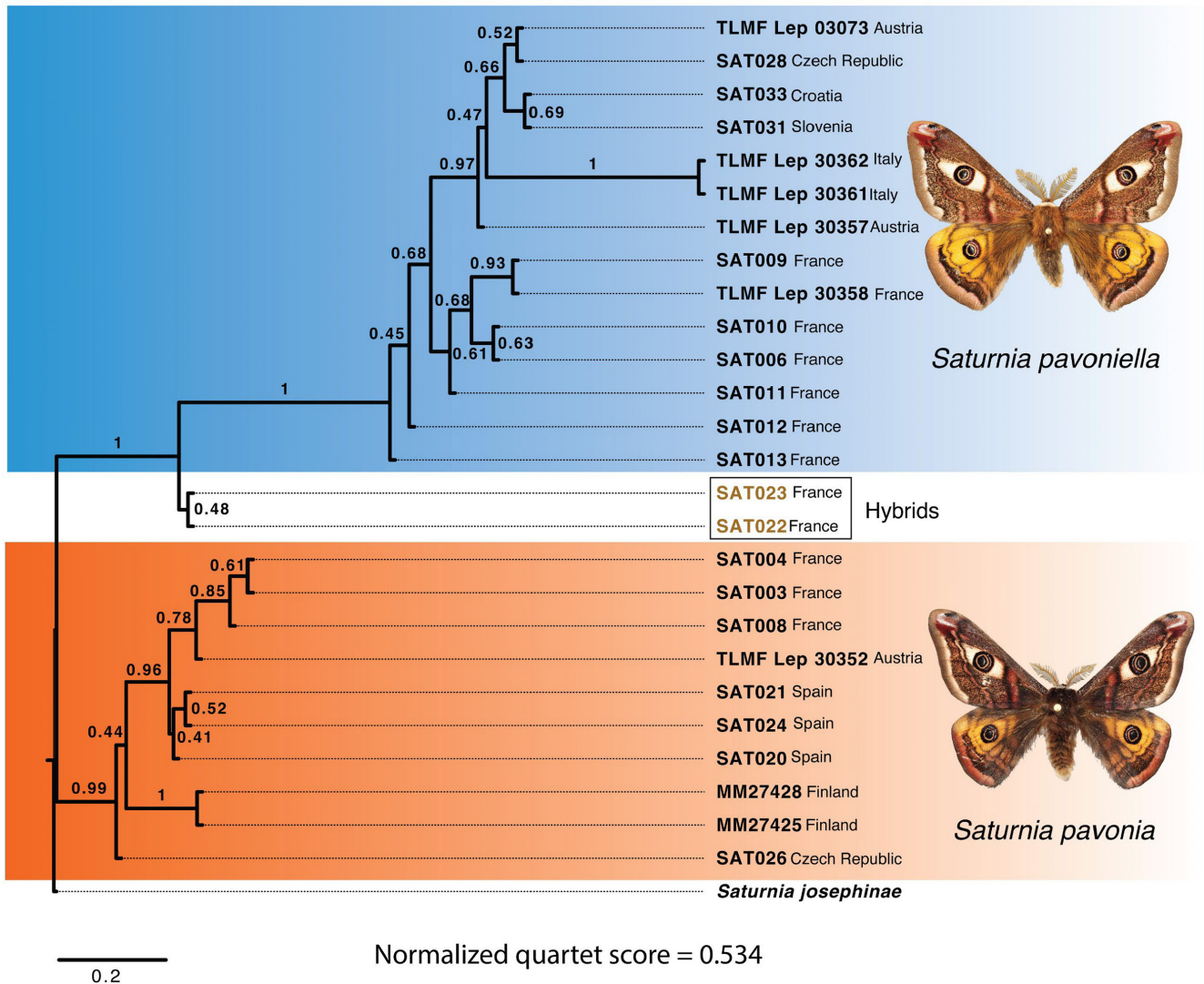


FIGURE 3 ASTRAL tree inferred for *Saturnia*. Numbers on the branches indicate quartet support values. Both specimens SAT022 (male) and SAT023 (female) indicated by the box are lab-reared F1 hybrids obtained from the following crossing: *pavoniella* female * *pavonia* male.

biological species concept (reproductive barrier present) and the phylogenetic species concept (reciprocal monophyly). The apparent lack of gene flow suggests that the taxa have developed barriers to gene flow, despite having a narrow geographical overlap. This is also evident from isolation by distance (IBD) analysis, where geographic distance was not found to be significant in explaining the observed genetic separation of the two taxa. Often, species that are a part of a parapatric distribution system show striking morphological and ecological similarities, but morphologically intermediate individuals may occur (Guiller et al., 2017; Slender et al., 2017; Tahami et al., 2021). In the present study system, *S. pavonia* and *S. pavoniella* are morphologically similar, although *S. pavoniella* is slightly larger. The differences between the wing patterns and the genitalia were noticed 20 years ago only, and despite being separated by Huemer and Nässig (2003), members of the two populations were commonly identified as varieties of a single species. A possibility for hybridization exists in and around the contact zone as pheromonal differentiation seems to be incomplete and females of both species attract

males of the other species. Natural hybridization has been reported by Huemer and Nässig (2003). Although hybridization occurs, it seems that the hybrids are sterile in natural populations as we could not see any signs of present introgression. Thus the postzygotic reproductive isolation seems to be complete.

Two evolutionary scenarios can explain the origin of parapatric mode of distribution: (1) parapatric speciation and (2) secondary contact of the populations that diverged in allopatry. While parapatric speciation is disputable and documented cases are few, there is evidence that under certain circumstances strong disruptive selection can promote speciation by driving adjacent populations to adapt to different ecological conditions and to differ through character displacement and reinforcement, eventually leading to reproductive isolation (Reifová et al., 2011). However, *S. pavonia* and *S. pavoniella* seem to share very much the same ecological niche and are attracted to the pheromones of "wrong" species, suggesting that they have not been in contact for a very long time as otherwise pheromonal differentiation would be expected. Thus, our observations are not

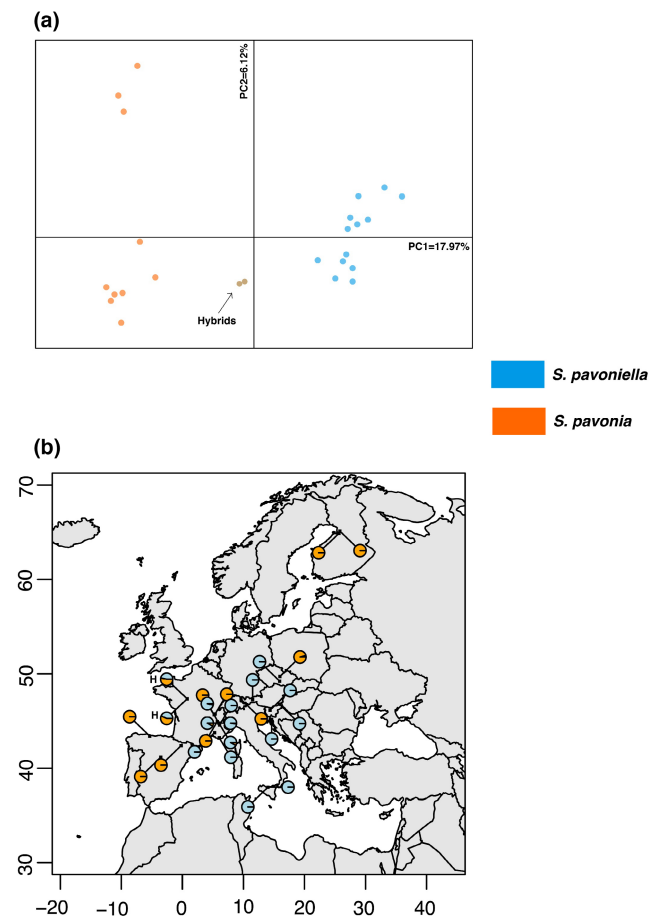


FIGURE 4 (a) PCA for *Saturnia* SNP dataset. (b) Pie charts based on membership coefficient matrix at $K=2$ mapped on geographic coordinates. Hybrid individuals are indicated by 'H'.

fully consistent with the scenario of speciation through ecological separation (due to adaptations to local habitats) under parapatry.

The two *Saturnia* populations presumably diverged during the last interglacial period in distinct refugia in allopatry. This isolation led to the appearance of postzygotic isolation by genetic drift (rather than selection), followed by the post-glacial expansion of ranges resulting in a secondary contact, as has been assumed to have taken place in other similar cases (Tahami et al., 2021). As body size is closely associated with the mating behaviour, the size difference between parapatric taxa could potentially result from character displacements in secondary contact periods (Zhang & Kubota, 2023). The possibility that the two *Saturnia* species evolved body size difference through this process can therefore not be ruled out.

In theory, the secondary contact of diverged populations might not only lead to a continuing speciation process through reinforcement (Pfennig, 2016), but also to the merging of the populations. As there is a mechanism at play that maintains a reproductive barrier between *S. pavonia* and *S. pavoniella*, as indicated by most of the analyses, the second scenario appears unlikely. Some evidence of past introgression was observed in our data, but it seems to have ceased and has not resulted in disruption of this barrier. In some cases, hybrid zones can also act as a 'genetic sink' that strongly restricts

gene flow between the species (Hafner et al., 1983; Yanchukov et al., 2006). Indeed, it has been suggested that the divergence in local allopatry is protected by hybrid zones as the taxa later become parapatric (Hewitt, 1988). To test this idea in the present study system, a more detailed study of samples from the contact zone would be required.

Parapatric species systems provide interesting models to study speciation in action. Events such as these take place at the interface of population genetics and phylogenetics levels for which reason analyses designed to address the question at both levels are potentially useful. Previously studied cases of parapatric species pairs or groups have revealed common patterns that suggested speciation not having reached completion makes taxonomic delimitation of parapatric taxa a very challenging and often largely arbitrary exercise, thus highlighting that speciation as a slow gradual process rather than a sudden event. From an evolutionary point of view, one might expect to observe systems at different stages of the speciation-population continuum, particularly when parapatry has resulted from a secondary contact of populations differentiated in allopatry, for example during glacial refugial periods.

4.3 | Potential of target enrichment in species delimitation

A number of recent studies have promoted the use of multi-locus genomic approaches such as ultraconserved elements (UCEs), ddRAD-sequencing, and target capture to aid species delimitation (Natusch et al., 2020; Rancilhac et al., 2019; Smith et al., 2014). Target capture is a cost-effective approach with greater sequencing depth and less data burden compared to whole genome sequencing (WGS) and whole exome sequencing (WES) (Bewicke-Copley et al., 2019). Evidently, the number of loci recovered in the present study was sufficient to clarify the taxonomic relationships between the two very closely related species. The percentage of missing data ranged from 18.56% to 35.08%, which is modest compared to those typically seen in RAD data sets (Lee et al., 2018). The amount of recovered data provided insights into evolutionary processes that are of high relevance in species delimitation, including admixture, presence of gene flow, and reciprocal monophyly. There is therefore little doubt that the used approach is powerful enough to address all the relevant evolutionary and other epistemological aspects needed for robust species delimitation.

This is a starting point to address perhaps even more difficult ontological and philosophical aspects of species delimitation. Because parapatric species systems may show a variety of patterns for example with regard to the levels of admixture and mode of divergence (or merging), it might not be possible to establish objective criteria for taxonomic delimitation of such taxa. Due to this inherent arbitrariness, we have before been calling for more consistency for delimitation of species under such complex evolutionary circumstances. This would serve taxonomic stability the best, not jeopardizing scientific validity. Needless to say, the two

Saturnia species this study has focused on do not represent the most complex case in the speciation continuity as their integrity as biological or phylogenetic species did not remain unclear. In case of ongoing but restricted gene flow, the cohesion (or lack of it) of the two lineages might provide a useful solution to species delimitation problem of admixing parapatric taxa (Templeton, 1989), although other feasible solutions may remain.

5 | CONCLUSIONS

We investigated the parapatric relationship of *S. pavonia* and *S. pavoniella* to elucidate genetic patterns and admixture under a complex evolutionary setting in which speciation of recently diverged populations seems not to have reached full completion. We further addressed the question if genomics tools such as target capture, which enables the accumulation of vast amounts of genetic data, could possibly provide better means to delimit species under complex circumstances, which, for instance, parapatric species pairs commonly represent. Our findings support the hypothesis that the two taxa are distinct species based on both biological and phylogenetic species concepts. We found evidence for mitonuclear discordance due to past introgression and incomplete lineage sorting. Our observations suggest that *S. pavonia* and *S. pavoniella* have experienced a secondary contact after a period of isolation. Periodic range expansions due to climate oscillations may have resulted in many secondary contacts of such diverged populations, which ultimately could lead to ecological differentiation through character displacement. Our study highlights the importance of understanding the evolutionary processes that shape the genetic diversity of organisms in parapatric relationships, which are likely to be common in nature. We found that species delimitation of parapatric species can at times be even relatively straightforward as we could not detect any ongoing introgression in *Saturnia*. However, the two reproductively isolated taxa have not yet completed speciation from an ecological standpoint as their sexual pheromones continue to attract heterospecific individuals and hybridizations occur. We assume that the lack of pheromonal differentiation plays a significant evolutionary role in maintaining parapatry in moths, as individuals dispersing into the area of the other species are likely to have low fitness. This situation creates evolutionary “pressure” for the two species to develop more reliable prezygotic isolation mechanisms, for example, through pheromonal character displacement. Although we lack direct evidence, we hypothesize that this process is underway. Conducting detailed studies focusing on the natural pheromone chemistry of the two species in and around the contact zone would provide valuable insights into this process.

AUTHOR CONTRIBUTIONS

MM, PH, ME and CLV conceived the study and designed the research project. ME participated in developing the target enrichment bait kit and molecular protocols. MM, PH and CLV collected

the specimens used in the study. CLV carried out the hybrid experiment and tested the synthetic sexual pheromone in the field. MJ and MK performed experiments, analysed the data and drafted the manuscript. MM and ME supervised the research. All authors commented and edited the draft and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank Laura Törmälä for her assistance in laboratory procedures and Vladislav Ivanov for constructive discussions regarding Dsuite analyses. Thanks are due to Till Tolasch for synthesizing and sending us the sex pheromone that was used to attract and collect specimens for this study. We thank Carlos Antonietty Adame, Yann Baillet, Philippe Bordet, Gregory Guicherd, Stanislav Gomboc, Zdenek Lastuvka, and Enrique Murria Beltran for sending *Saturnia* specimens from their respective home countries. The sequencing service was provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki. Additionally, we would like to thank CSC Finland for providing computational resources to carry out bioinformatics analyses. Finally, we would like to thank two anonymous reviewers whose comments improved the quality of the manuscript.

FUNDING INFORMATION

This study was supported by the Academy of Finland through Grant No. 314702 to MM. CLV was partly funded by the ERDF project InfoBioS (EX011185). MJ would like to thank the Finnish Cultural Foundation and the University of Oulu Scholarship Foundation for providing funding to continue her doctoral studies. MK was Co-funded by the European Union under the Marie Skłodowska-Curie Action I4WORLD 101081280. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency [REA]. Neither the European Union nor REA can be held responsible for them.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw target enrichment data are available at <https://www.ncbi.nlm.nih.gov/sra/PRJNA933196> with SRA Accession PRJNA933196. Alignments and other data are available on Zenodo Repository with the <https://doi.org/10.5281/zenodo.10021191>.

ORCID

Maria Khan  <https://orcid.org/0000-0003-2004-6391>

Mukta Joshi  <https://orcid.org/0000-0002-6785-8306>

Marianne Espeland  <https://orcid.org/0000-0002-6800-4783>

Peter Huemer  <https://orcid.org/0000-0002-0630-545X>

Carlos Lopez-Vaamonde  <https://orcid.org/0000-0003-2278-2368>

Marko Mutanen  <https://orcid.org/0000-0003-4464-6308>

REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Barley, A. J., White, J., Diesmos, A. C., & Brown, R. M. (2013). The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. *Evolution*, 67(12), 3556–3572. <https://doi.org/10.1111/evo.12219>
- Bergthorsson, U., Adams, K. L., Thomason, B., & Palmer, J. D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, 424(6945), Article 6945. <https://doi.org/10.1038/nature01743>
- Bewicke-Copley, F., Arjun Kumar, E., Palladino, G., Korfi, K., & Wang, J. (2019). Applications and analysis of targeted genomic sequencing in cancer studies. *Computational and Structural Biotechnology Journal*, 17, 1348–1359. <https://doi.org/10.1016/j.csbj.2019.10.004>
- Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A python package for genome-scale hybridization detection. *Systematic Biology*, 67(5), 821–829. <https://doi.org/10.1093/sysbio/syy023>
- Breinholz, J. W., Earl, C., Lemmon, A. R., Lemmon, E. M., Xiao, L., & Kawahara, A. Y. (2018). Resolving relationships among the mega-diverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Systematic Biology*, 67(1), 78–93. <https://doi.org/10.1093/sysbio/syx048>
- Bull, C. M. (1991). Ecology of parapatric distributions. *Annual Review of Ecology and Systematics*, 22, 19–36.
- Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, 22(17), 4369–4383. <https://doi.org/10.1111/mec.12413>
- Chambers, A. E., & Hillis, D. M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, 69(1), 184–193. <https://doi.org/10.1093/sysbio/syz042>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chernomor, O., Von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65(6), 997–1008. <https://doi.org/10.1093/sysbio/syw037>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- de Jong, R. (1987). Superspecies *Pyrgus malvae* (Lepidoptera: Hesperiiidae) in the East Mediterranean, with notes on phylogenetic and biological relationships. *Zoologische Mededelingen*, 61(34), 483–500.
- De Queiroz, K. (1999). The general lineage concept of species and the defining properties of the species category. In W. Robert (Ed.), *Species: New interdisciplinary essays* (pp. 49–89). MIT Press.
- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6), 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>
- deWaard, J. R., Ivanova, N. V., Hajibabaei, M., & Hebert, P. D. N. (2008). Assembling DNA barcodes. In C. C. Martin & C. C. Martin (Eds.), *Environmental genomics* (pp. 275–294). Humana Press. https://doi.org/10.1007/978-1-59745-548-0_15
- Dietz, L., Eberle, J., Mayer, C., Kukowka, S., Bohacz, C., Baur, H., Espeland, M., Huber, B. A., Hutter, C., Mengual, X., Peters, R. S., Vences, M., Wesener, T., Willmott, K., Misof, B., Niehuis, O., & Ahrens, D. (2022). Standardized nuclear markers improve and homogenize species delimitation in Metazoa. *Methods in Ecology and Evolution*, 00, 1–13. <https://doi.org/10.1111/2041-210X.14041>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- 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(4), 336–345. <https://doi.org/10.1016/j.tree.2019.12.003>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Ferrer Obiol, J., Herranz, J. M., Paris, J. R., Whiting, J. R., Rozas, J., Riutort, M., & González-Solis, J. (2023). Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. *Molecular Phylogenetics and Evolution*, 179, 107671. <https://doi.org/10.1016/j.ympev.2022.107671>
- Firmino, T. J., Jr., O'Neill, J. R., Itgen, M. W., Kihneman, T. A., Townsend, J. H., & Fujita, M. K. (2021). Delimitation despite discordance: Evaluating the species limits of a confounding species complex in the face of mitonuclear discordance. *Ecology and Evolution*, 11(18), 12739–12753. <https://doi.org/10.1002/ece3.8018>
- 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(5), 759–771. <https://doi.org/10.1093/sysbio/syw028>
- Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 397–423. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132421>
- Gaston, K. J., & Spicer, J. I. (2004). *Biodiversity: An introduction* (2nd ed.). Wiley-Blackwell. <https://www.wiley.com/en-us/Biodiversity%3A+An+Introduction%2C+2nd+Edition-p-9781405118576>
- Godinho, R., Llaneza, L., Blanco, J. C., Lopes, S., Álvares, F., García, E. J., Palacios, V., Cortés, Y., Tategón, J., & Ferrand, N. (2011). Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Molecular Ecology*, 20(24), 5154–5166. <https://doi.org/10.1111/j.1365-294X.2011.05345.x>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, 5, 184–186. <https://doi.org/10.1111/j.1471-8278>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), Article 7. <https://doi.org/10.1038/nbt.1883>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspina, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., ... Pääbo, S. (2010). A draft

- sequence of the Neandertal genome. *Science*, 328(5979), 710–722. <https://doi.org/10.1126/science.1188021>
- 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(2), 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(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hafner, J. C., Hafner, D. J., Patton, J. L., & Smith, M. F. (1983). Contact zones and the genetics of differentiation in the pocket gopher *Thomomys bottae* (Rodentia: Geomyidae). *Systematic Biology*, 32(1), 1–20. <https://doi.org/10.1093/sysbio/32.1.1>
- Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41(41), 95–98.
- Harmoinen, J., von Thaden, A., Aspi, J., Kvist, L., Cocchiararo, B., Jarausch, A., Gazzola, A., Sin, T., Lohi, H., Hytönen, M. K., Kojola, I., Stronen, A. V., Caniglia, R., Mattucci, F., Galaverni, M., Godinho, R., Ruiz-González, A., Randi, E., Muñoz-Fuentes, V., & Nowak, C. (2021). Reliable wolf-dog hybrid detection in Europe using a reduced SNP panel developed for non-invasively collected samples. *BMC Genomics*, 22(1), 473. <https://doi.org/10.1186/s12864-021-07761-5>
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105(S1), 795–809. <https://doi.org/10.1093/jhered/esu033>
- Hausdorf, B. (2011). Progress toward a general species concept. *Evolution*, 65(4), 923–931. <https://doi.org/10.1111/j.1558-5646.2011.01231.x>
- Hewitt, G. M. (1988). Hybrid zones—natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*, 3(7), 158–167. [https://doi.org/10.1016/0169-5347\(88\)90033-X](https://doi.org/10.1016/0169-5347(88)90033-X)
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3), 247–276. <https://doi.org/10.1006/bjil.1996.0035>
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405(6789), Article 6789. <https://doi.org/10.1038/35016000>
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1442), 183–195. <https://doi.org/10.1098/rstb.2003.1388>
- Hey, J. (2001). The mind of the species problem. *Trends in Ecology & Evolution*, 16(7), 326–329. [https://doi.org/10.1016/S0169-5347\(01\)02145-0](https://doi.org/10.1016/S0169-5347(01)02145-0)
- 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(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Huemer, P., & Näsigg, W. A. (2003). Der Pfauenspinner *Saturnia pavoniella* (SCOPOLI, 1763) sp. Rev. Im Gebiet der Ostalpen (Lepidoptera: Saturniidae). *Entomologische Zeitschrift*, 113(16), 180–190.
- Ivanov, V., Lee, K. M., & Mutanen, M. (2018). Mitonuclear discordance in wolf spiders: Genomic evidence for species integrity and introgression. *Molecular Ecology*, 27(7), 1681–1695. <https://doi.org/10.1111/mec.14564>
- Jackson, N. D., Carstens, B. C., Morales, A. E., & O'Meara, B. C. (2017). Species delimitation with gene flow. *Systematic Biology*, 66(5), 799–812. <https://doi.org/10.1093/sysbio/syw117>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Joshi, M., Espeland, M., Dincă, V., Vila, R., Tahami, M. S., Dietz, L., Mayer, C., Martin, S., Dapporto, L., & Mutanen, M. (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. <https://doi.org/10.1111/syen.12557>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jeremiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology and Systematics*, 24(1), 189–216. <https://doi.org/10.1146/annurev.ecolsys.24.1.189>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lanfear, R., Calcott, B., Kainer, D., Mayer, C., & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, 14(1), 1–14. <https://doi.org/10.1186/1471-2148-14-82>
- Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014). Species delimitation using genome-wide SNP data. *Systematic Biology*, 63(4), 534–542. <https://doi.org/10.1093/sysbio/syu018>
- Leaché, A. D., Zhu, T., Rannala, B., & Yang, Z. (2019). The spectre of too many species. *Systematic Biology*, 68(1), 168–181. <https://doi.org/10.1093/sysbio/syy051>
- Lee, K. M., Kivelä, S. M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N., & Mutanen, M. (2018). Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multi-locus Sanger data in a species-rich moth genus. *Systematic Biology*, 67(6), 925–939. <https://doi.org/10.1093/sysbio/syy029>
- Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution; International Journal of Organic Evolution*, 61(6), 1417–1438. <https://doi.org/10.1111/j.1558-5646.2007.00114.x>
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523–536. <https://doi.org/10.1093/sysbio/46.3.523>
- Mai, U., & Mirarab, S. (2018). TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, 19(5), 272. <https://doi.org/10.1186/s12864-018-4620-2>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite—Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595. <https://doi.org/10.1111/1755-0998.13265>
- Mallet, J. (2005). Speciation in the 21st century. *Heredity*, 95(1), Article 1. <https://doi.org/10.1038/sj.hdy.6800686>
- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, 7(2), 111–118. <https://doi.org/10.1038/nmeth.1419>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1), 209–220.
- Mayden, R. L. (1997). A hierarchy of species concepts: The denouement in the Saga of the species problem. In M. F. Claridge, H. A. Dawah, & M. R. Wilson (Eds.), *Species: The units of diversity* (pp. 381–423). Chapman & Hall.
- 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(3), 649–671. <https://doi.org/10.1111/syen.12481>
- Mazel, R. (2007). Contribution à l'étude du complexe *Saturnia pavonia* (L., 1758)—*S. pavoniella* (Scopoli, 1763) dans le sud de la France (Lepidoptera, Saturniidae). 3, 82–86.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Mirarab, S. (2019). Species tree estimation using ASTRAL: Practical considerations. *arXiv: Populations and Evolution*. <https://www.semanticscholar.org/paper/Species-Tree-Estimation-Using-ASTRAL%3A-Practical-Mirarab/abcf67ff57a284d258efdc3e1a389dc4f52cad0>
- 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(10), e47214. <https://doi.org/10.1371/journal.pone.0047214>
- Mutanen, M., Kivelä, S. M., Vos, R. A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., Huemer, P., Dincă, V., van Nieuwerkerken, E. J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaëns, T., Efetov, K. A., Hebert, P. D. N., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., ... 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. <https://doi.org/10.1093/sysbio/syw044>
- Natusch, D. J. D., Esquerré, D., Lyons, J. A., Hamidy, A., Lemmon, A. R., Moriarty Lemmon, E., Riyanto, A., Keogh, J. S., & Donnellan, S. (2020). Species delimitation and systematics of the green pythons (*Morelia viridis* complex) of melanesia and Australia. *Molecular Phylogenetics and Evolution*, 142, 106640. <https://doi.org/10.1016/j.ympev.2019.106640>
- Nosil, P. (2008). Ernst Mayr and the integration of geographic and ecological factors in speciation. *Biological Journal of the Linnean Society*, 95(1), 26–46. <https://doi.org/10.1111/j.1095-8312.2008.01091.x>
- Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R. S., Podsiadlowski, L., Vasilikopoulos, A., Zhou, X., Misof, B., & Niehuis, O. (2017). Orthograph: A versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, 18(1), 111. <https://doi.org/10.1186/s12859-017-1529-8>
- Pfennig, K. S. (2016). Reinforcement as an initiator of population divergence and speciation. *Current Zoology*, 62(2), 145–154. <https://doi.org/10.1093/cz/zow033>
- 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(2), 945–959.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Rancillac, L., Goudarzi, F., Gehara, M., Hemami, M. R., Elmer, K. R., Vences, M., & Steinfarz, S. (2019). Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. *Molecular Phylogenetics and Evolution*, 133, 189–197. <https://doi.org/10.1016/j.ympev.2019.01.003>
- Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., ... Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), Article 7327. <https://doi.org/10.1038/nature09710>
- Reifová, R., Reif, J., Antczak, M., & Nachman, M. W. (2011). Ecological character displacement in the face of gene flow: Evidence from two species of nightingales. *BMC Evolutionary Biology*, 11(1), 138. <https://doi.org/10.1186/1471-2148-11-138>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology*, 4(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Saino, N., Lorenzini, R., Fuscoti, G. E. A., Randi, 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(7), 605–613.
- Saino, N., & Villa, S. (1992). Pair composition and reproductive success across a hybrid zone of carrion crows and hooded crows. *The Auk*, 109(3), 543–555.
- Sayyari, E., & Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, 33(7), 1654–1668. <https://doi.org/10.1093/molbev/msw079>
- Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, 4, 11. <https://doi.org/10.1186/1742-9994-4-11>
- Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25), 16122–16127. <https://doi.org/10.1073/pnas.242585899>
- Slater, G., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6(1), 31. <https://doi.org/10.1186/1471-2105-6-31>
- 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(3), 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(1), 83–95. <https://doi.org/10.1093/sysbio/syt061>
- Sota, T., & Vogler, A. P. (2001). Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Systematic Biology*, 50(1), 39–59.
- Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, 16(8), Article 8. <https://doi.org/10.1038/nrg3962>
- Tahami, M. S., Dincă, V., Lee, K. M., Vila, R., Joshi, M., Heikkilä, M., Dapporto, L., Schmid, S., Huemer, P., & Mutanen, M. (2021). Genomics reveal admixture and unexpected patterns of diversity in a Parapatric pair of butterflies. *Genes*, 12, 2009. <https://doi.org/10.3390/genes12122009>
- Templeton, A. R. (1989). The meaning of species and speciation: A genetic perspective. In D. Otte & J. A. Endler (Eds.), *Speciation and its consequences* (pp. 3–27). Sinauer Associates.
- Tobias, J. A., Seddon, N., Spottiswoode, C. N., Pilgrim, J. D., Fishpool, L. D. C., & Collar, N. J. (2010). Quantitative criteria for species delimitation. *Ibis*, 152(4), 724–746. <https://doi.org/10.1111/j.1474-919X.2010.01051.x>
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>

- Yanchukov, A., Hofman, S., Szymura, J. M., Mezhzherin, S. V., Morozov-Leonov, S. Y., Barton, N. H., & Nürnbergger, B. (2006). Hybridization of *Bombina bombina* and *B. variegata* (anura, Discoglossidae) at a sharp ecotone in Western Ukraine: Comparisons across transects and over time. *Evolution*, 60(3), 583–600. <https://doi.org/10.1111/j.0014-3820.2006.tb01139.x>
- Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, 61(5), 854–865. <https://doi.org/10.1093/czoolo/61.5.854>
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19(Suppl 6), 15–30. <https://doi.org/10.1186/s12859-018-2129-y>
- Zhang, S.-N., & Kubota, K. (2023). Strongly exclusive distribution with putative character displacement in two flying stag beetles. *Evolution*, 77(1), 155–165. <https://doi.org/10.1093/evolut/qpac034>

SUPPORTING INFORMATION

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

How to cite this article: Khan, M., Joshi, M., Espeland, M., Huemer, P., Lopez-Vaamonde, C., & Mutanen, M. (2023). Patterns of speciation in a parapatric pair of *Saturnia* moths as revealed by target capture. *Molecular Ecology*, 00, 1–18. <https://doi.org/10.1111/mec.17194>