INTRODUCTION

Divergence in allopatry provides a simple null model of speciation (Mayr, 1947). Following geographic isolation and given enough time, reproductive isolation is inevitable as incompatibilities will eventually become fixed as a result of genetic drift and/or selection (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). Taxa that evolved partial reproductive isolation in allopatry may come into secondary contact as a result of range shifts and—depending on their degree of reproductive isolation and niche overlap—either form a contact zone or invade each other’s range (Barton, 1985; Pigot, 2013). If allopatric divergence dominates speciation, then local alpha diversity for a given clade cannot accrue until secondary sympatry is achieved (Weir & Price, 2011). Thus, the forces that facilitate or hamper secondary sympatry and the timescale over which this occurs have profound consequences both for speciation and for the spatial distribution of species diversity. While modern ranges only provide a snapshot of the dynamic history of range shifts, understanding the
extent to which current range overlap between closely related species can be explained by their speciation history and vice versa has been at the core of speciation research (Coyne & Orr, 2004).

The glacial cycles of the Pleistocene had a profound effect on current diversity of temperate ecosystems (Hewitt, 1996, 2001; Hofreiter, 2009). Populations of temperate taxa in Europe were isolated in ice-free refugia around the Mediterranean basin (Iberia, Italy, the Balkans and the larger Mediterranean islands) as glaciers encroached. The observation that the geographic ranges of many young taxa are restricted to individual glacial refugia in southern Europe (Dennis et al., 1991; Hewitt, 1996, 1999, 2011; Schmitt, 2007) suggests that this repeated separation into and expansion out of glacial refugia has played a major role in their origin. The availability of allozyme and mitochondrial (mt) data in the 80s and 90s has spurred an abundance of case studies on intra- and inter-specific diversity of European taxa including detailed investigations of hybrid zones in taxa ranging from fire-bellied toads (Kruuk et al., 1999), the house mouse (Boursot et al., 1996), grasshoppers (Barton and Butlin & Hewitt, 1985), to plants (Bacilieri et al., 1996) and marine mussels (Skibinski & Beardmore, 1979). The prevailing evidence from these studies is that genetic diversity within and in, many cases, divergence between species is structured by refugia (Dapporto et al., 2019; Hewitt, 1996; Schmitt, 2007).

### 1.1 When was divergence between sister species initiated?

While it is clear that the hybrid zones we observe today are secondary contacts that formed after the last glacial maximum and may have formed many times over throughout the Pleistocene, it is far from clear when divergence between the sister taxa involved was initiated. One possibility is that the Pleistocene glacial cycles initiated species divergence directly by separating populations into allopatric refugia (i.e. a ‘species pump’ sensu Haffer, 1969). Another possibility is that the initial divergence between sister species predates the Pleistocene, and so, any build-up of reproductive isolation during the Pleistocene (e.g. via the fixation of intrinsic incompatibilities and/or reinforcement) occurred in populations that were already partially diverged. If the Pleistocene species pump hypothesis is correct, we would expect sister species divergence times to be concentrated during or at the beginning of the mid-Pleistocene transition 0.8–1.2 million years ago (MYA), which marks the onset of continent-wide glacial cycling (Bishop et al., 2011). The idea that Pleistocene divergence acted as a species pump was first proposed in the context of American faunas (Avise et al., 1998; Bernatchez & Wilson, 1998; Haffer, 1969), but has dominated phylogeographic studies on European sister taxa (e.g. Habel et al., 2008; Hewitt, 1996, 2000; Schmitt, 2007; Schoville et al., 2012). In contrast, other studies including some of the early work on European contact zones (Barton & Hewitt, 1985; Butlin & Hewitt, 1985) conclude that the taxa involved in such secondary contacts may substantially pre-date the Pleistocene (Abbott et al., 2000; Hewitt, 1996; Klicka & Zink, 1997; Spooner & Ritchie, 2006). Similarly, Pleistocene climate forcing is insufficient in explaining divergence in an Amazonian butterfly suture zone (Dasmahapatra et al., 2010). Thus, it remains unclear to what extent divergence between sister taxa was initiated by ‘Pleistocene species pump’ dynamics or has an older, deeper origin?

A corollary for the hypothesis of allopatric speciation in different refugia is that range overlap is secondary. Since species can more easily invade each others’ ranges once sufficient premating barriers and ecological differentiation have developed, we would expect species pairs with overlapping ranges to be older overall than those without range overlap, all else being equal (Coyne & Orr, 2004).

Support for this prediction comes from comparative studies showing that the proportion of range overlap (degree of sympathy (Chester & Zink, 1994)) is positively (albeit weakly) correlated with genetic divergence (Barraclough & Vogler, 2000; Pigot & Tobias, 2013). However, a recent study in Chorthippus grasshoppers shows that subspecies that hybridize across contact zones can be older than currently sympatric species (Nolen et al., 2020).

### 1.2 Mitonuclear discordance

Age estimates for recently diverged taxa have largely relied on single-locus phylogenies and ignored incomplete lineage sorting. Hewitt (2011) summarizes age estimates for European hybrid-zones taxa including mammals, insects, amphibians and reptiles, which range from hundreds of thousands to several million years ago. However, given that these estimates are based on different markers and calibrations, the extent to which glacial cycles have initiated speciation events remains unknown. Estimates based on mitochondrial (mt) data are particularly unreliable for at least three reasons. First, the mutation rate of mtDNA is highly erratic (Galtier et al., 2009). Second, given the stochasticity of coalescence, the ancestry of a single locus (however well resolved) is a very poor measure of species divergence. In the absence of gene flow, divergence at a single locus may substantially pre-date the onset of species divergence, while it may be much more recent in the presence of gene flow (Knowles & Carstens, 2007; Wang & Hey, 2010). Mitonuclear discordance in both directions has been found in a large number of animal systems (Toews & Brelsford, 2012) including several closely related species of European butterflies (Dincă et al., 2019; Hinojosa et al., 2019; Wiemers et al., 2010). Finally, mtDNA does not evolve neutrally since transmission of mitochondria is completely linked to maternal inheritance of endosymbionts such as Wolbachia and Spiroplasma and, in organisms with Z/W sex determination, of the W chromosome. Thus, mt diversity and divergence may be driven largely by selective sweeps (including introgression sweeps) rather than neutral gene flow and genetic drift (Galtier et al., 2009; Hurst & Jiggins, 2005; Jiggins, 2003; Martin et al., 2020).

### 1.3 European butterflies as a model group

Testing whether climate-induced Pleistocene range shifts have triggered speciation or patterned older splits between species requires
replication both at the level of genetic loci and at the level of speciation events. Although we can now generate WGS data for any species, there are surprisingly few reliable estimates for the onset of divergence between European sister species and such estimates are lacking even for well-studied contact zone taxa (but see Nolen et al., 2020; Nürnberg et al., 2016).

Lepidoptera are arguably the best-studied arthropod family: European butterflies provide a unique opportunity to investigate divergence and speciation processes comparatively (Dapporto et al., 2019). Near-complete information on geographic ranges and key life history traits (e.g. voltinism and host plant range) is available (Kudrna, 2019; Tolman & Lewington, 2013). Additionally, the taxonomy of all 496 European species is well resolved (Wiemers et al., 2018) and a complete, multilocus phylogeny of all European taxa exists (Dapporto et al., 2019). This, combined with extensive DNA barcode reference libraries (Dapporto et al., 2019; Dincă et al., 2021), facilitates the identification of species (especially in the case of cryptic taxa) and provides extensive sampling of sister species pairs, many of which abut at narrow contact zones (Dennis et al., 1991; Platania et al., 2020; Vodă et al., 2015) (Figure 1). Secondary contact zones have been described in detail for several European taxa, including *Spialia orbifer* and *S. sertorius* (Lorkovic, 1973), the Italian *Pontia* hybrid zone (Porter et al., 1997) and the contacts between *Iphiclides podalirius* and *I. feisthamelii* and between *Melanargia galathea* and *M. lachesis* along the Pyrenees (Habel et al., 2017; Wohlfahrt, 1996, Gaunet et al., 2019).

Here, we use European butterflies as a model system to investigate to what extent the divergence times between sister species in this group are concentrated in the Pleistocene, as predicted by the Pleistocene species pump hypothesis, and test how well recent sister species fit a null model of divergence in allopatry. Although European butterflies have been studied intensively, with few exceptions (see Talla et al., 2017), robust estimates of divergence required for any systematic comparison of speciation are lacking. We generate RNA-seq data for 18 sister species pairs and ask the following specific questions:

(i) Has speciation been initiated during the Pleistocene, as envisaged by the species pump hypothesis, or did the glacial cycles pattern pre-existing, older subdivisions?
(ii) Are sister species pairs that form contact zones younger than pairs that overlap in range?
(iii) Is there evidence for gene flow between contact zone species?
(iv) How strongly correlated are mitochondrial and nuclear divergence and do contact zone pairs show increased mitonuclear discordance?

**Figure 1** Nine of the 18 sister species pairs of butterfly in which we quantified genome-wide divergence meet at contact zones in southern Europe. In the left group, from left to right across northern Iberia are *Satyrium, Pseudophilotes, Melanargia* and *Iphiclides*. In the centre group, from bottom to top across the Alps are *Pontia, Euchloe, Pyrgus*, and *Zerynthia*. Finally, on the right across the Balkans is the genus *Spialia*. 
2 | METHODS

2.1 | Sampling and molecular work

We identified true sister species pairs in the European butterfly phylogeny (Wiemers et al., 2018; Dapporto et al., 2019). Species pairs involving island and mountain endemics were excluded, as these cannot achieve secondary sympatry. We also did not consider species pairs that are unlikely to have originated in Europe, for example sister pairs involving North American taxa. Following these criteria, we sampled 18 sister species pairs (Table 1). Our sampling includes 7.3% of European butterfly species (Wiemers et al., 2018) and almost all ‘good’ butterfly sister species pairs in Europe (Descimon & Mallet, 2009).

Field sampling was conducted over multiple seasons (2016–2019) at several locations across southern and central Europe (Portugal, Spain, France, Hungary, and Romania) targeting known glacial refugia (and avoiding localities close to known contact zones) whenever possible. Samples were hand-netted in the field, flash-frozen in a liquid nitrogen dry shipper (Voyageur 12) and stored at −70°C shortly after capture (wings were retained for identification).

Specimen identifications were confirmed for 14 species that are difficult to identify based on morphology (morphological characters are subtle and/or internal) but for which COI barcodes are diagnostic of morphoID using LepF/R primers (Hajibabaei et al., 2006) and existing reference databases (Dincă et al., 2021). We were unable to obtain fresh material for Erebia euryale and E. ligea, and Fabriciana adippe and F. niobe (two remaining sister pairs meeting our sampling criteria).

RNA extractions were prepared by dividing individuals bilaterally (including all parts of the body: head, thorax and abdomen) and using one side. RNA was extracted following a hybrid protocol: samples were homogenized in TRIzol, and RNA was eluted using the Purelink RNA Purification Kit protocol after a DNA digestion step. Extracted RNA was submitted to Edinburgh Genomics to generate automated TruSeq-stranded mRNA-seq libraries. Libraries were sequenced on an Illumina NovaSeq platform using 100-bp paired-end reads after poly-A selection. For each species, where possible, we generated RNA-seq data for two samples, one male and one female from different localities. Transcriptome data for 66 samples (across 38 species) were generated and analysed previously by Mackintosh et al., (2019). Of these, 26 samples from 13 species are included in the present analysis (Table S1).

2.2 | Generating transcriptome assemblies

Reads were processed following the pipeline developed by Mackintosh et al., (2019). Reads were trimmed and checked for quality using FastQc v0.11.8 (Andrews et al., 2010) both before and after trimming with FastP v0.20.0 (Chen et al., 2018) using MultiQC v1.7 (Ewels et al., 2016) to visualize the results. Trimmed reads were assembled into de novo transcriptomes using Trinity v2.8.5 (Grabherr et al., 2011), pooling data sets by species. Transcriptome completeness was assessed using BUSCO v5 (Simão et al., 2015) on full transcriptome assemblies using the lep-idopteraodb10 database, with scores ranging from 84 to 96% (Table S2). Transcripts were processed with Transdecoder v5.5 (Haas et al., 2016) and retained based on BLAST (Camacho et al., 2009) and HMMER (Finn et al., 2011) homology search results. Read pairs from each sample were mapped against the respective species transcriptome, composed of the longest isoform of each complete protein-coding transcript, using BWA MEM (Li, 2013). Coverage at mapped sites was determined using GATK CallableLoci v3.5 (McKenna et al., 2010). Sites with at least 10-fold coverage and a minimum MAPQ of 1 in each sample were considered suitable for variant calling. Callable loci were intersected between individuals using BEDTools v2.28 (Quinlan & Hall, 2010), and variants were called using FreeBayes v1.3.1 (Garrison & Marth, 2012) and filtered for unbalanced SNPs and missing genotypes (RPL ≥ 1 & RPR ≥ 1 & SAF ≥ 1 & SAR ≥ 1 & N_MISSING = 0) using BCFTools filter v0.1.19 (Li, 2011).

To generate comparable data sets across all samples, Orthofinder v2.3.3 (Emms & Kelly, 2015) was used to cluster proteins into orthogroups. Orthogroups were labelled single-copy orthologues (SCOs) if one protein of each taxon was present. Genus single-copy orthologues (GSCOs) were diagnosed based on the presence of single-copy proteins within the focal pair. Protein sequences from each orthogroup were used to align corresponding DNA sequences using Translatorx v12.0 (Abascal et al., 2010).

Data were generated for 36 species (18 sister pairs) from five families. For 17 pairs, data were generated from 665 SCOs from high-quality transcriptomes (BUSCO scores >90%). For the pair of Zerynthia species (one of which, Zerynthia polyxena, was sampled as a larva), GSCOs (5000 orthologues) were used to avoid restricting the SCOS for other pairs. With the exception of the Zerynthia pair, all analyses are based on SCO to enforce consistent comparisons across pairs. While the SCO data set is much smaller than the pair GSCO data sets and likely enriched for conserved and highly expressed genes, this has very little impact on estimates of divergence and diversity at fourfold degenerate (4D) sites, as these are highly correlated (Figure S1 and Mackintosh et al., 2019).

2.3 | Estimating gene and population divergence

For each species pair, we calculated the average pairwise gene divergence $d_{xy}$ at fourfold degenerate (4D) sites using sequence alignments for one or two diploid samples from each species. This calculation is implemented in the script orthoderiver.py (www.github.com/samebdon/orthoderiver).

Species split times were estimated using two different approaches. First, we used a simple rescaling of mean genetic divergence and diversity to obtain a lower bound of the species divergence time. Assuming a simple null model of divergence without gene flow, neutrality and an infinite site mutation model, mean neutral divergence $d_n = d_{xy} - \pi$ (Nei & Li, 1979) is directly proportional to species divergence time $T = \frac{d_n}{2\mu}$. Here, $\mu$ is the de novo mutation rate per generation (per base).
<table>
<thead>
<tr>
<th>Sister 1</th>
<th>( \pi )</th>
<th>Gen ~1</th>
<th>Sister 2</th>
<th>( \pi )</th>
<th>Gen ~1</th>
<th>( d_{xy} )</th>
<th>( d_s )</th>
<th>Split time (MYA)</th>
<th>( F_{st} )</th>
<th>Degree of sympathy</th>
<th>Contact zone</th>
<th>Known to hybridize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brenthis daphne</td>
<td>0.0046</td>
<td>1</td>
<td>B. ino†</td>
<td>0.0094</td>
<td>1</td>
<td>0.0246</td>
<td>0.0176</td>
<td>3.04 (1.60, 6.89)</td>
<td>0.716</td>
<td>0.74</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Colias alfacariensis†</td>
<td>0.0243</td>
<td>2–3</td>
<td>C. hyale†</td>
<td>0.0211</td>
<td>2–3</td>
<td>0.0387</td>
<td>0.0159</td>
<td>0.92 (0.48, 2.05)</td>
<td>0.412</td>
<td>0.70</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Euchloe ausonia</td>
<td>0.0250</td>
<td>2</td>
<td>E. crameri†</td>
<td>0.0352</td>
<td>2</td>
<td>0.0715</td>
<td>0.0416</td>
<td>3.58 (1.89, 7.99)</td>
<td>0.582</td>
<td>0.00</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Gonepteryx cleopatra</td>
<td>0.0304</td>
<td>1</td>
<td>G. rhamni</td>
<td>0.0156</td>
<td>1</td>
<td>0.0448</td>
<td>0.0316</td>
<td>5.48 (2.89, 12.2)</td>
<td>0.705</td>
<td>0.97</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Iphiclides feithomelian§</td>
<td>0.0079</td>
<td>1–3</td>
<td>I. podalirius†</td>
<td>0.0052</td>
<td>1–3</td>
<td>0.0275</td>
<td>0.0204</td>
<td>1.20 (0.63, 2.68)</td>
<td>0.747</td>
<td>0.00</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lasiommatata megera</td>
<td>0.0385</td>
<td>2–3</td>
<td>L. petropolitana§</td>
<td>0.0065</td>
<td>1</td>
<td>0.0543</td>
<td>0.0316</td>
<td>2.75 (1.45, 6.13)</td>
<td>0.586</td>
<td>0.43</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Leptidea reali§</td>
<td>0.0077</td>
<td>1–2</td>
<td>L. sinapis§</td>
<td>0.0093</td>
<td>1–3</td>
<td>0.0153</td>
<td>0.0068</td>
<td>0.47 (0.25, 1.04)</td>
<td>0.444</td>
<td>1.00</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Melanargia galathea†,‡</td>
<td>0.0152</td>
<td>1</td>
<td>M. lachesis§</td>
<td>0.0145</td>
<td>1</td>
<td>0.0389</td>
<td>0.0227</td>
<td>4.14 (2.18, 9.24)</td>
<td>0.597</td>
<td>0.20</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pieris manni‡</td>
<td>0.0100</td>
<td>1</td>
<td>P. rapae§</td>
<td>0.0198</td>
<td>3–4</td>
<td>0.0678</td>
<td>0.0525</td>
<td>2.61 (1.37, 5.81)</td>
<td>0.779</td>
<td>1.00</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Polyommatus eros‡</td>
<td>0.0104</td>
<td>1</td>
<td>P. icarus§</td>
<td>0.0174</td>
<td>1–3</td>
<td>0.0529</td>
<td>0.0382</td>
<td>3.37 (1.78, 7.51)</td>
<td>0.724</td>
<td>1.00</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pontia daplidice†,‡</td>
<td>0.0063</td>
<td>3</td>
<td>P. edusa</td>
<td>0.0159</td>
<td>3</td>
<td>0.0516</td>
<td>0.0401</td>
<td>2.33 (1.22, 5.19)</td>
<td>0.777</td>
<td>0.00</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pseudophilotes baton</td>
<td>0.0080</td>
<td>1–2</td>
<td>P. panopes</td>
<td>0.0131</td>
<td>1</td>
<td>0.0276</td>
<td>0.0171</td>
<td>1.97 (1.03, 4.39)</td>
<td>0.619</td>
<td>0.00</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Pyrgus malvae†</td>
<td>0.0164</td>
<td>1–2</td>
<td>P. malvoides‡</td>
<td>0.0176</td>
<td>1–2</td>
<td>0.0362</td>
<td>0.0192</td>
<td>1.66 (0.87, 3.70)</td>
<td>0.531</td>
<td>0.04</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Satyrirum esculi§</td>
<td>0.0076</td>
<td>1</td>
<td>S. iialis</td>
<td>0.0036</td>
<td>1</td>
<td>0.0432</td>
<td>0.0378</td>
<td>6.49 (3.42, 14.5)</td>
<td>0.869</td>
<td>0.06</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Satyris actea†</td>
<td>0.0261</td>
<td>1</td>
<td>S. ferala‡</td>
<td>0.0074</td>
<td>1</td>
<td>0.0663</td>
<td>0.0493</td>
<td>8.53 (4.50, 19.0)</td>
<td>0.745</td>
<td>0.26</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Spialia orbifer†,‡</td>
<td>0.0331</td>
<td>2</td>
<td>S. sertorius</td>
<td>0.0418</td>
<td>2</td>
<td>0.0671</td>
<td>0.0292</td>
<td>2.55 (1.35, 5.69)</td>
<td>0.438</td>
<td>0.12</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Thymelicus actea‡</td>
<td>0.0154</td>
<td>2</td>
<td>T. sylvestris‡</td>
<td>0.0208</td>
<td>1</td>
<td>0.0848</td>
<td>0.0671</td>
<td>7.66 (4.04, 17.1)</td>
<td>0.790</td>
<td>0.99</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Zerynthia cassandra</td>
<td>0.0033</td>
<td>1</td>
<td>Z. polyxena</td>
<td>0.0032</td>
<td>1</td>
<td>0.0432</td>
<td>0.0279</td>
<td>4.82 (2.54, 10.7)</td>
<td>0.895</td>
<td>0.00</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Gen ~1 is the number of generations per year. Samples from species marked with a † were barcoded to confirm correct identification. Species where Wolbachia presence was confirmed by Duplouy and Hornett (2018) and Hinojosa et al., (2019) are marked with a ‡.
We assumed $\mu = 2.9 \times 10^{-7}$, an estimate of the spontaneous mutation rate obtained from parent–offspring trios of South American Heliconius melpomene butterflies (Keightley, Pinharanda, et al., 2014). Since both violations of the mutation model (back-mutations) and the demographic model (gene flow) reduce $d_{xy}$ this time estimate is a lower bound of the true species divergence time. We converted estimates of species divergence time ($T$) into years ($\tau$) using the mean generation time of each pair (Table 1). Information on generation times was compiled from Collins Butterfly Guide (Tolman & Lewington, 2013) (Table 1). For species in which generation times vary with latitude, we assumed the minimum generation time of the southern part of the range. This is a reasonable long-term average, given that European glacial refugia are located around the Mediterranean.

Second, we estimated species split times using a multilocus coalescent approach. We considered the distribution of pairwise differences in blocks of a fixed length of 4D sites to fit an isolation-with-migration (IM) model and a nested history of strict divergence to each species pair. In the absence of recombination within blocks, the distribution of pairwise differences has been derived analytically (Lohse et al., 2011; Wilkinson-Herbots, 2012). We obtained maximum-likelihood parameter estimates under both models and used likelihood-ratio tests for model comparisons in Mathematica v11.3 (File S1). The block length for each pair was chosen to give an average of three pairwise differences between sister species per block.

However, given the high rate of recombination (relative to mutation) in butterflies (Martin et al., 2016, 2019) and the substantial span of 4D blocks, we expect multilocus inference (assuming no within-block recombination) to be biased. In particular, recombination is known to lead to (upwardly) biased estimates of divergence time (Wall, 2003).

Given this and other limitations (see Discussion), we will focus on the more simple and robust estimates of species divergence based on $d_{xy}$ throughout.

As an additional test for gene flow, we compared the observed distributions of pairwise differences with analytic expectations under a model of strict divergence without gene flow given the estimates of $T$ and ancestral $N_e$ obtained from $d_{xy}$ and mean $\alpha$ (which cannot be affected by recombination).

Thus, in the absence of gene flow, we would expect the empirical distribution of pairwise differences to be narrower than the analytic expectation (Wall, 2003) due to recombination. In contrast, wider-than-expected distributions are indicative of post-divergence gene flow. We re-sampled (without replacement) 10,000 data sets of equal size as the observed data from the expected distribution of each species and tested whether the likelihood of the observed distribution of pairwise differences falls within the distribution of likelihoods expected under a null model of strict divergence.

### 2.4 Estimating range size and overlap

Geographic ranges were quantified as follows: we obtained occurrence data over Europe for all focal species with a resolution of 60’ latitude and 30’ longitude by critically revising the data from the Distribution Atlas of European Butterflies and Skippers (Kudrna et al., 2011) and by adding data from Roger Vila’s collection stored at Institut de Biologia Evolutiva (Barcelona) (Figures S2–S4). To calculate range overlap, we applied the biodiencypt function (Platania et al., 2020) of the recluster R package (Dapporto et al., 2013). This function computes alpha hull with a given concavity ($\alpha$) and evaluates the area of overlap among pairs of species. We used $\alpha = 2$ and $\alpha = 3$ for species with discontinuous and continuous distributions in Europe, respectively. We quantified the range overlap of each species pair and calculated the degree of sympathy as:

$$\text{Sympathy} = \frac{\text{Overlap}_{A,B}}{\min(\text{Area}_{A,B})},$$

that is the fraction of the smaller range involved in the overlap. In the following, we consider sister pairs with a degree of sympathy ≤0.2 contact zone pairs and those with a degree of sympathy >0.2 sympatric.

Based on this, we classified nine pairs as contact zone taxa. However, since there are only two species pairs with intermediate levels of sympathy (>0.2 and <0.7), our comparisons of contact zone and sympatric pairs are robust to a wide range of thresholds.

### 2.5 Mitochondrial diversity and divergence

Sequence alignments for the COI barcode locus were obtained from the BOLD database (Ratnasingham & Hebert, 2007) for all 18 sister species pairs. These sequences, together with associated information, are publicly available in the data set DS-EUGENMAP (dx.doi.org/10.5883/DS-EUGENMAP) on BOLD at www.boldsystems.org and were originally produced by Dincă et al. (2021). For each species, we included all available sequence records from Europe (ranging from 21 in E. crameri to 429 in L. sinapis (Table S1)). Mean pairwise diversity ($\pi$) within species and divergence ($d_{xy}$) across all sites were computed using DnaSP (Librado & Rozas, 2009).

We obtained the average gene divergence time for each pair from the multilocus-calibrated phylogeny of European butterflies of Wiemers et al., (2020) as half of patristic distances calculated with distTIPS function of the adephylo R package (Jombart & Dray, 2010). The correlation between our estimates of species divergences and these node ages was explored with standardized major axis (SMA) regression, using the ‘sma’ function of the ‘smatr’ R package. SMA estimates slope and intercept and tests whether slope differs from one.

### 3 RESULTS

#### 3.1 Most European butterfly sister species predate the Pleistocene

Mean gene divergence ($d_{xy}$) at 4D sites between sister species ranged from 0.015 to 0.085, with a mean of 0.047 (Table 1, Figure 2) across
the 18 pairs. We contrasted sympatric and allopatric samples in one genus where it was possible to confirm that sampling location has negligible impact on estimates of divergence (M. galathea (Romania): M. lachesis (Spain) \(d_s = 0.039\), Spain: Spain = 0.038). Divergence between samples from the same localities typically varied by 0.0005.

Species divergence times obtained from \(d_s\) at fourfold degenerate sites (4D) ranged from 0.47 (0.2, 1.0) (Leptidea) to 8.5 (4.5, 19) (Satyrs) MYA, with a mean of 3.8 MYA (Figure 2). The distribution of divergence estimates does not suggest there is synchronicity in divergence independent of absolute values. Even though these point estimates are lower bounds of species divergence (see Discussion), they not only substantially predate the mid-Pleistocene transition (15 out of 18 pairs) but also, in the majority of cases (11 out of 18 pairs), are older than the entire Quaternary period ≈2.6 MY (Table 1). Of the seven taxa with Pleistocene \(r\) estimates, three fall in the early Pleistocene: Pseudophilotes (1.97 (1.0, 4.4) MYA), Pontia (2.33 (1.2, 5.2) MYA) and Spialia (2.55 (1.3, 5.7) MYA). However, when accounting for the (known) uncertainty in the mutation rate estimate we used as a molecular clock (Keightley, Pinharanda, et al., 2014), we can rule out a divergence time more recent than the mid-Pleistocene transition or the Quaternary period for 13 and four sister species pairs, respectively (Figure 2).

We find that estimates of sister species divergence based on the distribution of pairwise differences are highly correlated (Pearson's correlation coefficient \(= 0.79\)) with the estimates based on mean \(d_s\) (Figure 5S). The IM model fit significantly better than a strict divergence model for two species pairs (Pontia and Colias), and, as expected, \(T\) estimates under the IM model for these two species pairs are older (3.30 vs 0.919 and 5.08 vs 2.33 MYA, respectively) than estimates based on \(d_s\) (Table S3). However, we find that sampling blocks of a fixed length resulted in a consistent downward bias in \(d_{xy}\) of on average 10%. This is likely a result of selecting for long and likely conserved transcripts. In the light of this, and given that blocks based on 4D sites violate other key assumptions of multilocus inference, in particular, no recombination within loci and known phase, we caution against over-interpreting these model-based estimates (see Discussion) and focus on the simpler and more robust estimates of sister species divergence based on \(d_s\) throughout.

3.2 | Sister pairs that form contact zones are not significantly younger than sympatric pairs

There are two reasons to expect species pairs that form contact zones to be younger than sympatric pairs: first, if speciation under a null model of divergence in allopatry is initiated by periods of vicariance, the formation of a contact zone (parapatry) represents an earlier stage in the transition to complete reproductive isolation and substantial range overlap (sympathy) (Coyne & Orr, 2004). Second, any gene flow across contact zones would reduce \(d_s\) and hence our estimate of species divergence. While the nine pairs that form contact zones (degree of sympathy \(<0.2\)) have a lower net divergence (\(d_s = 0.0287, SD = 0.0093\)) than the nine sympatric pairs (degree of sympathy \(>0.2\), \(d_s = 0.0347, SD = 0.0195\); Table 1), this difference is not significant (\(t = -0.82999, df = 11.478, p = 0.210\)). Additionally, we find no relationship between the degree of sympatry and \(d_s\) (\(t = 0.723, df = 16, p = 0.480\)). Similarly, we may expect pairs that are still able to form hybrids (i.e. for which F1s have been observed in the wild) to be younger than those that do not. However, contrary to this expectation, we again find no significant difference in \(d_s\) between pairs which do and do not hybridize (\(d_s = 0.0293\) and 0.0329, respectively, \(t = -0.582, df = 15.861, p = 0.284\)).

3.3 | Evidence for recent gene flow in some contact zone pairs

The empirical distribution of pairwise differences deviates significantly from the expectation under a strict divergence model in a majority of species pairs (12 out of 18) (Figure 3 & S6). Of these, eight pairs have narrower distributions than expected, compatible with recombination within blocks, and four pairs have wider distributions than expected, compatible with post-divergence gene flow (Pseudophilotes, Pontia, Iplicides and Zerynthia). While the eight pairs with narrower distributions are equally split between contact and sympatric pairs, all four taxa with wider distributions are contact zone pairs (Figure 3). However, given the limited number of pairs overall, this difference between contact zones and sympatric pairs is not significant (Fisher's exact test, \(p = 0.0901\)).

3.4 | Pervasive mitonuclear discordance in contact zone species pairs

Our estimates of species divergence are based on average net divergence (\(d_{xy}\)) across many hundreds of genes and are robust to how orthologues are filtered (Figure S1). Given that previous studies on European butterflies have been largely based on mitochondrial (mt) phylogenies, an obvious question is to what extent mt divergence is correlated with mean nuclear divergence. We find that both \(d_s\) and \(d_{xy}\) at COI are positively but only weakly correlated with mean nuclear divergence (Figure 4). The correlation is weaker for \(d_s\) than for \(d_{xy}\) (\(R^2 = 0.27\) and 0.31, respectively), which is unsurprisingly given that mitochondrial diversity (and hence \(d_s\)) is both inherently random and may be affected by selective sweeps. Comparing the relation between \(mt\) and nuclear \(d_s\) between contact zone and sympatric pairs, we find a shallower slope for contact zone pairs (0.29 compared to 0.99; Figure 4). This difference, although not significant (\(p = 0.09\)), appears to be driven by the reduced \(mt\) diversity in contact zone compared with sympatric pairs (mean \(\kappa = 0.0030, SD = 0.0014\) and \(\kappa = 0.0047, SD = 0.0031\), respectively; \(t = 1.5763, df = 11.324, p = 0.0712\)). This suggests that \(mt\) diversity may be more strongly affected by selective sweeps in contact zone species than in sympatric pairs. We find no corresponding signal for any difference in nuclear
diversity between contact zones and sympatric pairs ($t = -0.0139$, $df = 31.539$, $p = 0.506$) and, in general, no correlation between nuclear and mt diversity (Figure S7 and Mackintosh et al., 2019). Our estimates for the lower bound of sister species divergence differ substantially from the ages of the corresponding nodes in the Wiemers et al., 2020 phylogeny for individual pairs (Figure S8). This is unsurprising given that the latter are largely informed by mtDNA data (Figure S9). However, perhaps surprisingly (given the difference in calibration, data and inference approach), our estimates are not consistently older or younger than the node ages of Wiemers et al., 2020 ($t_{paired} = -1.105$, $df = 17$, $p = 0.285$). A standardized major axis regression shows a significant relationship ($R^2 = 0.3657$, $p = 0.00780$), a slope (1.377) not different from one ($r = 0.3786$, $p = 0.121$) and an intercept ($-0.5750$) not different from zero (Figure S8).

3.5 | Genetic diversity does not correlate with relative range size

Genetic diversity at 4D sites within all 36 species ranged from 0.32% to 4.2% with a mean of 1.5%. Given the $H. \text{mel}pomene$ mutation rate of $\mu = 2.9 \times 10^{-9}$ (Keightley, Pinharanda, et al., 2014), these correspond to effective population sizes ranging from 280,000 to 3,600,000 with a mean of 1,300,000 (assuming $\theta = 4N_e \mu$). Mackintosh et al. (2019) tested whether neutral genetic diversity across European butterflies correlates with geographic range and found no significant relation across 38 taxa. Our sampling of species pairs allows for a simpler, alternative test of the potential relationship between diversity and range size using sister-clade comparisons, which is less sensitive to potential phylogenetic correlates and uncertainty in current range estimates. If diversity is a function of range size, we expect the species in a pair with the larger range to have higher genetic diversity than the species with the smaller range. We indeed find a difference in the expected direction, 0.0167 ($SD = 0.0114$) vs 0.0139 ($SD = 0.00865$), although the effect of relative range size is not significant ($t_{paired} = 1.127$, $df = 17$, $p = 0.138$; Figure S10).

4 | DISCUSSION

We quantify and compare genome-wide divergence across 18 sister species pairs of European butterfly. Simple estimates for the onset of species divergence based on net gene divergence ($d_J$) and a direct mutation rate estimate for butterflies suggest that the majority of pairs have diverged before the onset of the major Pleistocene glacial
FIGURE 3 The distribution of pairwise differences, \( S \) in blocks of a fixed length of 4D sites in contact (upper box) and sympatric (lower box) pairs. The observed distribution in single-copy orthologues is shown in orange, and the expectation under a history of strict divergence (estimated from \( \pi \) and \( d_{j} \)), in grey. Pairs that show wider-than-expected distributions are marked with an asterisk (*), and species that show narrower-than-expected distributions are marked with a plus (+).
cycles. Our results support the notion that the modern contact zones are secondary between species that began to diverge much earlier, in the Pliocene or early Pleistocene. Thus, even though the current ranges of many taxon pairs reflect glacial refugia, their initial divergence during the Pliocene or early Pleistocene is unlikely to have been triggered by repeated cycles of range connectivity and vicariance into refugia, as envisaged by the species pump hypothesis and phylogeographic studies based on mt and allozyme data (e.g. Habel et al., 2008; Lai & Pullin, 2004; Schmitt, 2007; Todisco et al., 2010; Zinetti et al., 2013) because substantial glaciation across continents did not develop (Bishop et al., 2011) until the ‘mid-Pleistocene transition’ from 0.8 to 1.2 MYA and the shift from 41,000- to 100,000-year glacial cycles. Given the antiquity of most sister species, it is perhaps unsurprising that we do not find any relationship between current range overlap and the time since divergence. Specifically, species pairs that form contact zones are not significantly younger than pairs that broadly overlap in range. However, we do find that strong signals of post-divergence gene flow are restricted to contact zone pairs. It is likely that the absence of sympatric pairs with significant gene flow reflects a simple survivorship bias: any incipient species pairs with significant gene flow might have already collapsed. Likewise, we are more likely to observe old contact zones pairs that have survived repeated glacial cycles.

Our finding that mt divergence between sister species is only weakly correlated with mean nuclear divergence and the possibility that net mt divergence may be greater for contact zone than sympatric species pairs as a result of reduced genetic diversity (note that the differences in $d_\mu$ and mean $\mu$ between contact zone and sympatric pairs are marginally non-significant, $p = 0.09$ and 0.07, respectively) could suggest that contact zone species may be subject to more frequent selective sweeps linked to mitochondria. Such sweeps may be acting on mt variation directly or indirectly through maternally inherited genomes or chromosomes (e.g. Wolbachia (Jiggins, 2003) and the W chromosome) and have been documented in a number of Lepidopteran systems (Ganet et al., 2019; Graham & Wilson, 2012; Kodandaramaiah et al., 2013; Martin et al., 2020; Ritter et al., 2013) (see Table 1 for species in this study with confirmed Wolbachia presence). Our results raise the intriguing possibility that such sweeps could play a role in the build-up of reproductive isolation (Giordano et al., 1997; Rokas, 2000; Shoemaker et al., 1999).

4.1 | Sources of dating uncertainty and bias

Since we have assumed the simplest possible demographic null model of species divergence without gene flow using $d_\mu$, our estimates of divergence between sister species based on $d_\mu$ should be interpreted as lower bounds. Any gene flow between sister species would reduce $d_\mu$ and species divergence estimates both by decreasing $d_\mu$ and by potentially increasing $\mu$ (in the recipient species).

Calibrating absolute split times involves assumptions about both the generation time and the mutation rate. We have assumed that the mutation rate is the same (per generation) across all species pairs, irrespective of their generation time, and applied a directed laboratory estimate of the per generation mutation rate from the tropical butterfly H. melpomene. While there is good evidence for a generation time effect on mutation rates in invertebrates (Thomas et al., 2010), our assumption of a simple linear relationship between generation time and sequence divergence may be overly simplistic. In particular, if temperate European species, which have longer average generation times than H. melpomene, have a higher per generation mutation rate, we would have overestimated the age of sister species. In contrast, given that generation time varies between populations, species and likely through time, our use of the average minimum generation time (within each pair) as a proxy for the long-term generation time is conservative: assuming longer average generation times would yield even older estimates species divergence. Likewise, while our assumption that 4D sites evolve under strict neutrality may be unrealistic, it is conservative with respect to our inference of old sister species divergence. For example, assuming that a fraction of 0.22 of 4D sites is under strong selective constraint due to codon usage bias (Lawrie et al., 2013) would result in underestimation of $T$ by 22%. Given these uncertainties in calibration, our absolute time estimates should be interpreted with caution until direct mutation rate estimates for temperate butterflies are available. However, in the absence of information on mutation rate heterogeneity across Lepidoptera, our main conclusion that divergence between most sister species of European butterflies predates the Pleistocene would still hold if mutation rates were higher by a factor of two. Given that the direct estimate of the de novo mutation rate in H. melpomene is similar to spontaneous mutation rate estimates for other insects (Keightley et al., 2014), this seems extremely unlikely. While our split time estimates may be surprising in the light of previous phylogeographic studies on European butterflies based on mt diversity (e.g. Habel et al., 2008; Lai & Pullin, 2004; Schmitt, 2007; Todisco et al., 2010; Zinetti et al., 2013), our divergence estimates for Leptidea rea I and L. sinapis, the youngest and only pair for which divergence has been estimated based on genome-wide data before, is lower than previous estimates (Talla et al., 2017).

4.2 | Glacial cycling and the Messinian salinity crisis

Taking our estimates of species splits at face value, the divergence of eleven species pairs predate the onset of Pleistocene glacial cycling >2.6 MYA (Gibbard & Head, 2009). This is not compatible with the idea that, overall, recent speciation events in European butterflies were initiated by the range shifts into and out of glacial refugia during the Pleistocene. However, our age estimates do of course not rule out that Pleistocene range shifts and vicariance may have played an important role in completing speciation processes, for example through reinforcement and/or the evolution of intrinsic incompatibilities.
A major geographic event that could have initiated species divergence in Europe before the onset of Pleistocene glacial cycling is the Messinian salinity crisis (MSC) ≈ 6 MYA during which the Mediterranean greatly reduced in size (Krijgsman et al., 1999). As a consequence, Europe and Africa were connected across the strait of Gibraltar until the Zanclean flood when the Atlantic reconnected to the re-expanding Mediterranean sea. It is plausible that this created a strong dispersal barrier for many species that previously had continuous distributions around the Mediterranean basin and may have initiated divergence into the east and west European/Mediterranean sister taxa. While the MSC has been considered a plausible trigger of species divergence in amphibians (Nürnberger et al., 2016) and reptiles (Kaliontzopoulou et al., 2011) and butterflies in the Melitaea radiation (Leveneu et al., 2009), this remains of course speculation.

4.3 | Do European butterfly species fall within the grey zone of speciation?

Roux et al., (2016) conducted a comparative analysis of divergence and gene flow across 61 pairs of sister taxa and found that pairs with net synonymous divergence of >2% rarely show evidence for ongoing gene flow. In contrast, taxa with $d_s$ between 0.5% and 2% may show some evidence for ongoing gene flow and ambiguous species status, suggesting that speciation may be incomplete. While our five youngest pairs (Brenthis, Colias, Leptidea, Pseudophilotes and Pyrgus) fall in this ‘grey zone of speciation’, we only find evidence for gene flow in one (Pseudophilotes). In contrast, we find a clear gene flow signal in three more diverged pairs: Iphiclides, $d_s = 2.09$%; Zerynthia, $d_s = 2.79$%; and Pontia, $d_s = 4.05$%. However, as we have focused sampling on ‘good species’ sensu Mallet (Descimon & Mallet, 2009) we are missing the recent (intraspecific) end of the continuum of divergence described by Roux et al., (2016). It will be interesting to test to what extent intraspecific divergence times between different refugial populations of butterflies are concentrated in the mid-Pleistocene, a pattern that has been found for other herbivorous insect and their parasitoids (Bunnefeld et al., 2018), and quantify the overlap between inter- and intraspecific divergence times. Nevertheless, our contrasting finding of both gene flow signals in old contact zone pairs (e.g. Pontia) and no evidence for gene flow (and complete sympatry) in the youngest pair (Leptidea) suggests that the ‘grey zone of speciation’ may be very wide indeed for European butterflies.

4.4 | Outlook

Given the challenges of demographic inference from transcriptome data (in particular the high relative recombination rate in butterflies), we have deliberately resisted the temptation to overinterpret models of demographic history. Our goal was instead to establish robust and comparable lower bounds for the age of butterfly sister species in Europe. Being based on mean divergence at 4D sites, these lower
bounds for species ages make minimal assumptions and are unaffected by recombination. Likewise, we have decided to focus on a simple and conservative diagnostic for introgression.

Resolving these speciation processes in greater detail will require examination of whole-genome data from larger samples under realistic models of speciation history. Fitting explicit models of speciation to whole-genome sequence data, ideally including both selection and gene flow will undoubtedly refine estimates for the onset of divergence between these species pairs and overcome many of the biases inherent in basing such inferences on transcriptome data. Perhaps even more importantly, it would allow us to quantify the likely endpoints (if present) of speciation processes. While it is straightforward to determine lower bounds for the onset of divergence under simple null models that assume no gene flow, as we have done here, estimating upper bounds of species divergence in the presence of gene flow is a much harder inference problem. As pointed out by Barton and Hewitt (1985), the initial time of divergence may be unknowable given that post-divergence gene flow eventually erases all information about this parameter. Although current and historic levels of gene flow between European butterfly sister species remain to be determined, our results already suggest that their speciation histories are older and potentially slower than had been assumed by previous phylogeographic studies based on mt data. It will be fascinating to imagine the evolutionary forces that drive both this general pattern and its exceptions, in particular, the selection responsible for the origin of very young but complete (in terms of reproductive isolation) cryptic species such as Leptidea (Talla et al., 2019) and the recently discovered Spialia rosea (Hernández-Roldán et al., 2016).

ACKNOWLEDGEMENTS

This work was supported by an ERC starting grant (ModelGenomLand). SE is supported by an EastBio studentship from the British Biological Sciences Research Council (BBSRC). KL is supported by a fellowship from the Natural Environment Research Council (NERC, NE/L011522/1). LD is supported by a scholarship from the British Biological Sciences Research Council (BBSRC) David Phillips Fellowship (BB/N020146/1). We thank Alex Mackintosh and Simon Martin for helpful comments on an earlier draft, Richard Lewington for permission to reproduce his butterfly illustrations, Andrés de la Fila for help in the molecular laboratory and Edinburgh Genomics for sequencing.

AUTHOR CONTRIBUTIONS

SE, DRL, RV and KL conceived the study. SE performed molecular labwork and analysed the data. DRL, LD and KL provided tools for analyses. AH, LD and RV contributed samples. KL, DRL and MGR supervised the project. SE, KL and LD wrote the paper with input from all coauthors.

ETHICAL APPROVAL

Field sampling of butterflies was conducted in compliance with the School of Biological Sciences Ethics Committee at the University of Edinburgh and the ERC ethics review procedure. Permits for field sampling were obtained from the Generalitat de Catalunya (SF/639), the Gobierno de Aragon (INAGA/500201/24/2018/0614 to Karl Wotton) and the Gobierno del Principado de Asturias (014252). The samples for Zoys cassandra from Elba were collected after permission from the Italian Ministero dell’ Ambiente e della Tutela del Territorio e del Mare (Prot. 0012493/ PN M 24/06/2015).

DATA AVAILABILITY STATEMENT

Read data are available from the ENA at PRJEB43082. Sequence alignments for the COI barcode locus were obtained from the dataset DS-EUGENMAP (dx.doi.org/10.5883/DS-EUGENMAP) on BOLD at www.boldsystems.org and were originally produced by Dincă et al., (2021). The script used for calculating diversity and divergence is available at https://github.com/samebdon/ortho diver/

ORCID

Sam Ebdon https://orcid.org/0000-0002-9321-4686
Dominik R. Laetsch https://orcid.org/0000-0001-7887-0186
Leonardo Dappporto https://orcid.org/0000-0001-7129-4526
Alexander Hayward https://orcid.org/0000-0001-7413-718X
Michael G. Ritchie https://orcid.org/0000-0001-7913-8675
Vlad Dincă https://orcid.org/0000-0003-1791-2148
Roger Vila https://orcid.org/0000-0002-2447-4388
Konrad Lohe https://orcid.org/0000-0001-9918-058X

REFERENCES

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N.


Trends in Ecology & Evolution


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.


EBDON ET AL. 3589