ORIGINAL ARTICLE

Contrasting genomic and phenotypic outcomes of hybridization between pairs of mimetic butterfly taxa across a suture zone

Jérémy Gauthier[1,2] (iD) | Donna Lisa de Silva[3] | Zachariah Gompert[4] (iD) |
Annabel Whibley[5] (iD) | Céline Houssin[3] | Yann Le Poul[3,6] (iD) | Melanie McClure[3] (iD) |
Claire Lemaitre[1] (iD) | Fabrice Legeai[1] (iD) | James Mallet[7] (iD) | Marianne Elias[3] (iD)

[1]Inria, CNRS, IRISA, University Rennes, Rennes, France

[2]Geneva Natural History Museum, Geneva, Switzerland

[3]Institut de Systématique, Évolution, Biodiversité, CNRS, MNHN, EPHE, Sorbonne Université, Université des Antilles, Paris, France

[4]Department of Biology, Utah State University, Logan, UT, USA

[5]School of Biological Sciences, University of Auckland, Auckland, New Zealand

[6]Fakultat für Biologie, Biozentrum, Ludwig-Maximilians Universität München, Planegg-Martinsried, Germany

[7]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

Correspondence
Jérémy Gauthier, Geneva Natural History Museum, Geneva, Switzerland.
Email: mr.gauthier.jeremy@gmail.com

Funding information
Agence Nationale de la Recherche, Grant/Award Number: CLEARWING ANR-16-CE02-0012 and SPECREP ANR-14-CE02-0011; Human Frontier Science Program, Grant/Award Number: RGP0014/2016

Abstract

Hybrid zones, whereby divergent lineages come into contact and eventually hybridize, can provide insights on the mechanisms involved in population differentiation and reproductive isolation, and ultimately speciation. Suture zones offer the opportunity to compare these processes across multiple species. In this paper we use reduced-complexity genomic data to compare the genetic and phenotypic structure and hybridization patterns of two mimetic butterfly species, *Ithomia salapia* and *Oleria onega* (Nymphalidae: Ithomiini), each consisting of a pair of lineages differentiated for their wing colour pattern and that come into contact in the Andean foothills of Peru. Despite similarities in their life history, we highlight major differences, both at the genomic and phenotypic level, between the two species. These differences include the presence of hybrids, variations in wing phenotype, and genomic patterns of introgression and differentiation. In *I. salapia*, the two lineages appear to hybridize only rarely, whereas in *O. onega* the hybrids are not only more common, but also genetically and phenotypically more variable. We also detected loci statistically associated with wing colour pattern variation, but in both species these loci were not over-represented among the candidate barrier loci, suggesting that traits other than wing colour pattern may be important for reproductive isolation. Our results contrast with the genomic patterns observed between hybridizing lineages in the mimetic *Heliconius* butterflies, and call for a broader investigation into the genomics of speciation in Ithomiini - the largest radiation of mimetic butterflies.

KEYWORDS
admixture mapping, differentiation, hybridization, introgression, Ithomiini, mimicry

1 | INTRODUCTION

Speciation is the ultimate process responsible for the considerable biological diversity observed on Earth. Hybrid zones, whereby divergent lineages come into contact and potentially hybridize, can provide insights into the mechanisms involved in population differentiation and reproductive isolation, and ultimately speciation

(Barton & Hewitt, 1989; Ravinet et al., 2017; Safran & Nosil, 2012). When hybrid zones span an environmental transition, populations across the hybrid zone diverge not only because of genetic drift but also due to local adaptation to different environments. Over time, drift and selection can lead to the emergence of barriers to gene flow that increase reproductive isolation, resulting in heterogeneous patterns of differentiation and introgression across the genome (Barton

& Bengtsson, 1986; Ravinet et al., 2017; Safran & Nosil, 2012). Genomic regions with low rates of introgression are more likely to be associated with divergent selection and reproductive isolation (Gompert & Buerkle, 2009; Heliconius Genome Consortium, 2012; Jay et al., 2018; Mallet, 2005; Mallet & Barton, 1989). Assessing the genetic structure of hybrid zones can therefore shed light on the evolutionary processes at play during the early stages of speciation by revealing the number and distribution of loci presenting deviant patterns of differentiation and introgression compared to the genome-wide average (Bierne, Welch, Loire, Bonhomme, & David, 2011). Additionally, while genetic mapping of adaptive traits has classically relied on controlled crosses, which cannot be performed in many organisms, hybrid zones enable the application of admixture mapping approaches that take advantage of natural mixing and recombination to investigate the genetic basis underlying adaptive phenotypic variation (Buerkle & Lexer, 2008; Gompert & Buerkle, 2013; Pallares, Harr, Turner, & Tautz, 2014).

Many studies have focused on hybrid zones to unravel the processes generating local adaptation (e.g., Jones et al., 2012; Larson, Andrés, Bogdanowicz, & Harrison, 2013; Soria-Carrasco et al., 2014) and reproductive isolation (e.g., Christe et al., 2016; Teeter et al., 2008) and have revealed different patterns (Barton & Hewitt, 1985, 1989; Gompert & Buerkle, 2009; Kronforst et al., 2013; Rieseberg, Whitton, & Gardner, 1999; Teeter et al., 2010; Via & Hawthorne, 2002). These differences may stem from differences in the organisms studied, but may also stem from differences in the environmental conditions faced by populations of these organisms, such that comparative interpretations are rather limited. To investigate the repeatability of genetic and phenotypic differentiation, reproductive isolation and introgression patterns across incipient species diverging under similar environmental conditions are needed. Suture zones are areas where multiple recently diverged pairs of taxa come into contact and hybridize (Remington, 1968), and typically span a sharp environmental gradient or a dispersal barrier (Dasmahapatra, Lamas, Simpson, & Mallet, 2010; Endler, 1977; Moritz et al., 2009). With replicated pairs of divergent and hybridizing lineages, suture zones offer an exceptional opportunity to compare levels and patterns of hybridization and reproductive isolation in relation to genomic and phenotypic divergence in a common environmental setting (Moritz et al., 2009; Nosil, Funk, & Ortiz-Barrientos, 2009; Rissler & Smith, 2010).

Müllerian mimicry in butterflies, in which multiple defended species locally converge on warning wing colour patterns and form mimicry "rings" (Bates, 1862; Muller, 1879), provides an excellent system to unravel the mechanisms underlying adaptation and speciation. Two large neotropical mimetic butterfly tribes, Heliconiini and Ithomiini (Nymphalidae, 77 and 393 species, respectively) are particularly well suited. Heliconiine and ithomiine species typically comprise multiple geographical subspecies that differ in colour patterns (Brown, Sheppard, & Turner, 1974). Because warning colour patterns in mimetic butterflies are under strong positive frequency-dependent selection locally (Kapan, 2001; Mallet & Barton, 1989), divergent mimetic subspecies are often separated by narrow hybrid zones

maintained by migration-selection balance (Mallet & Barton, 1989). Mimicry generates postzygotic reproductive isolation via higher predation on intermediate-patterned, nonmimetic hybrids (Merrill et al., 2012) and prezygotic reproductive isolation if there is also assortative mating for colour pattern among subspecies (Jiggins, Naisbit, Coe, & Mallet, 2001; McClure et al., 2019; Merrill et al., 2011, 2012). Mimicry is therefore a strong ecological driver of speciation, and is believed to have triggered the diversification of large radiations of heliconiine and ithomiine butterflies (Jiggins et al., 2001; Kozak et al., 2015). Studies of genetic differentiation and the basis of colour pattern variation in mimetic butterflies have almost exclusively focused on heliconiine butterflies (particularly the genus *Heliconius*), where a few major-effect genes (dubbed the mimicry "toolkit" [Joron et al., 2006]) have been found to control wing pattern variation (Martin et al., 2012; Mazo-Vargas et al., 2017; Nadeau et al., 2016; Reed et al., 2011; Westerman et al., 2018) and to be highly differentiated across hybrid zones, while the rest of the genome seems highly permeable (Nadeau et al., 2014). By contrast, because of practical limitations (difficulties in maintaining captive populations and making controlled crosses), much less is known about population genetics of Ithomiini species (but see Dasmahapatra et al., 2010; McClure & Elias, 2016; McClure et al., 2019), let alone of the genetics of wing pattern variation. Yet, Ithomiini numerically dominate forest communities of day-flying Lepidoptera and are believed to be some of the main drivers of both Müllerian and Batesian mimicry (whereby palatable species mimic unpalatable aposematic species) among Lepidoptera in the Neotropics (Bates, 1862; Beccaloni, 1997; Muller, 1879). Because of the ecological importance of Ithomiini, shedding light on how populations are structured across hybrid zones and elucidating the genetic basis of wing pattern variation would significantly advance our understanding of adaptation and speciation in mimetic butterflies.

The foothills of the Andes in the region of Tarapoto are transitional between lowland rainforest and mid-elevation mountain forest. This area is a major suture zone for a range of organisms (Roberts et al., 2006; Smith et al., 2014; Weir, 2006) including heliconiine and ithomiine mimetic butterflies (Dasmahapatra et al., 2010; Nadeau et al., 2014; Whinnett et al., 2005) that harbour divergent wing colour patterns across the suture zone. Here we take advantage of the Tarapoto suture zone to assess the patterns of genomic and phenotypic divergence and infer the extent of reproductive isolation across the hybrid zone for two widespread ithomiine species, *Ithomia salapia* and *Oleria onega*, that each harbour divergent wing colour patterns across the Tarapoto suture zone. The two species have broadly similar life histories (e.g., forest habitats, mimicry, Solanaceae hostplants) and although they are somewhat differently distributed throughout the Neotropics, in the area of Tarapoto their populations have similar distributions. In this area, *I. salapia* comprises two subspecies: *I. salapia aquinia*, on the Amazonian side, which has a transparent yellow colour pattern surrounded by an orange and black line with small white dots; and *I. salapia derasa*, on the Andean side above about 500 m alt., has a similar yellow pattern but surrounded by a thick black line with large white dots (Figure 1).
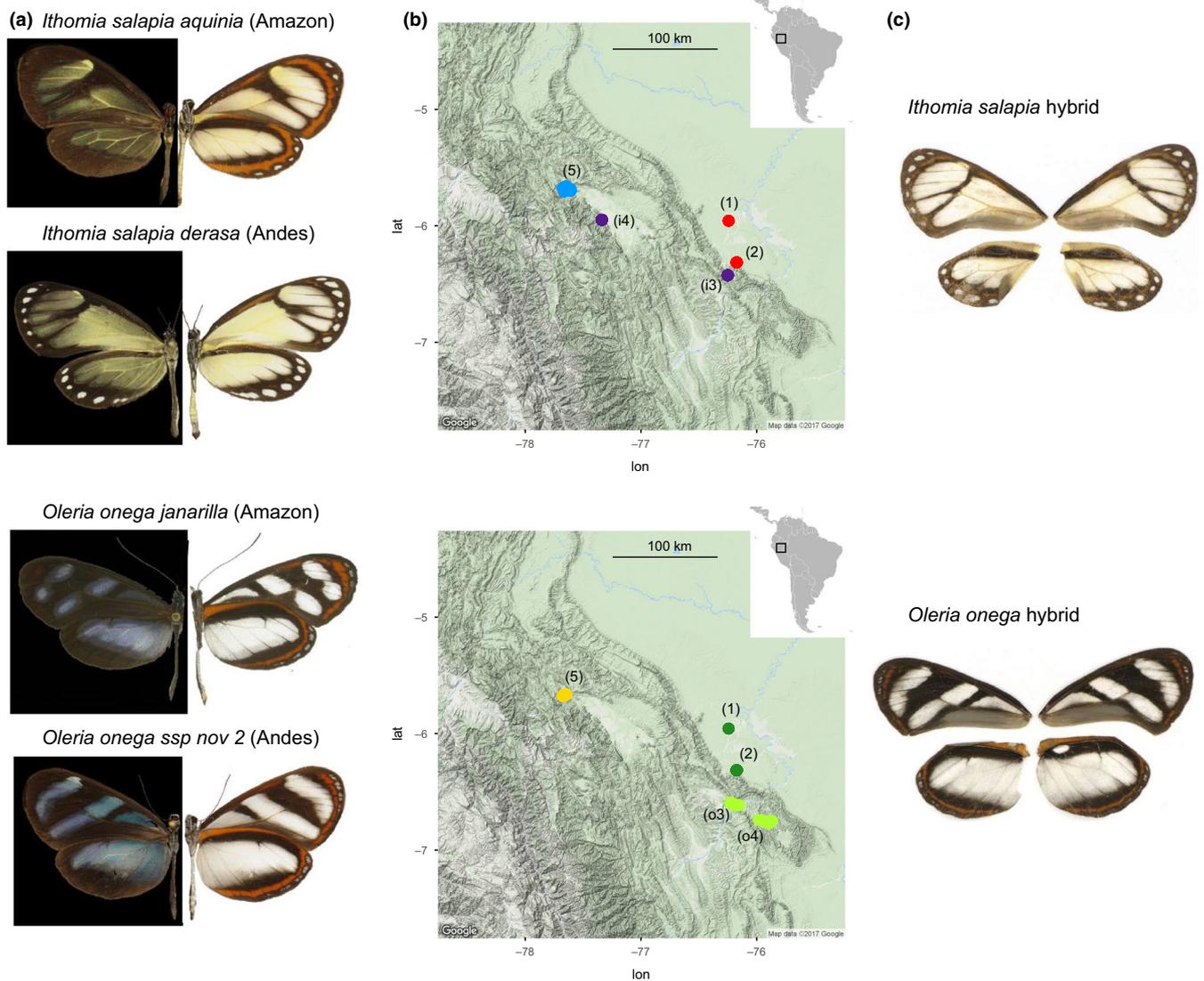
FIGURE 1 The study organisms and sites studied in N.E. Peru. (a) Photos of representative specimens from each population of the two studied species (dorsal side shown against a dark background to highlight transparency and ventral side shown against a white background to highlight colour pattern). (b) Sampling sites for *I. salapia* populations (top) with Amazonian sites in red (1) Km-26 Yurimaguas-Tarapoto and (2) San Miguel de Achinamiza, the Andean sites in blue (5) Puente Aguas Verdes and sites within the hybrid zone in purple (i3) Km-42 Tarapoto-Yurimaguas and (i4) La Florida. For *O. onega* populations (bottom), Amazonian sites are in green (1) Km-26 Yurimaguas-Tarapoto and (2) San Miguel de Achinamiza, Andean sites are in yellow (5) Puente Serranoyacu and the sites in the hybrid zone are in apple green (o3) Shapaja-Chazuta (o4) from Quebrada Yanayacu to Laguna del Mundo Perdido. Colour codes are conserved for all other figures. (c) Photographs of putative hybrid specimens with intermediate colour patterns (ventral side). Photograph credits: Céline Houssin

In addition to being found on the Amazonian side of the Escalera mountains, *I. salapia aquinia* is also found on the other side of the mountains, in the lowlands of the Río Mayo valley near Tarapoto.

On the Andean side (including in the lowlands of the Río Mayo) *O. onega* ssp. nov. 2 has translucent white wings with black patterning that splits the apical part of the forewing into two white 'windows', while in the Amazonian subspecies *O. onega janarilla* the forewing black patterning nearly splits the apical part of the forewing into four windows (Figure 1).

Thus, whereas *I. salapia* has a strictly altitude-based distribution of populations, *O. onega* has populations that differ geographically East-West, independently of altitude. For each species, the two

subspecies belong to distinct mimicry rings, which they numerically dominate. Therefore, *I. salapia* and *O. onega* probably play a major role in divergence and maintenance of their respective mimicry rings. Individuals with intermediate patterns are sometimes found in the contact zone, suggesting that occasional interlineage hybrids are produced.

In this paper we carry out population genomic analyses and quantify wing colour pattern variation in samples from Andean, Amazonian and intermediate populations of *I. salapia* and *O. onega* to address the following questions: (a) what are the overall patterns of differentiation, admixture and introgression between subspecies for each taxon pair? (b) To what extent does introgression vary across the genome,

and are genetic regions associated with colour pattern among those that exhibit higher differentiation and reduced introgression? (c) How repeatable are the patterns of differentiation, introgression and genotype-phenotype association across the two taxon pairs?

2 | MATERIALS AND METHODS

2.1 | Sampling

Sampling was performed in five study sites in the region of Tarapoto in Peru (Figure 1a; details in Table S1: gives each site, GPS coordinates and the number of each sex of *I. salapia* and *O. onega* sampled, before and after filtering). A total of 112 *I. salapia* specimens were sampled from five localities, and 149 *O. onega* specimens were sampled in five localities.

2.2 | Genotyping by sequencing (GBS)

DNA was extracted from one-quarter of thorax of each individual using the Qiagen DNeasy blood and tissue kit, following the manufacturer protocol. We generated reduced genomic complexity libraries for each specimen using a GBS (genotyping by sequencing) approach (Gompert et al., 2012; Parchman et al., 2012). Briefly, genomic DNA was digested with the restriction endonucleases EcoRI and MseI and resulting fragments were ligated to double-stranded adaptor oligonucleotides. These adaptors consisted of the Illumina sequencing priming sites followed barcodes that allow for the identification of sequences for each individual. These barcodes allowed us to multiplex all individuals into one library. Sequencing of the library was completed by the National Center for Genome Research (Santa Fe, NM, USA) on an Illumina HiSeq platform; 100 base pair single-end sequencing reads were generated.

2.3 | SNP calling

First, sequencing primers were removed, sequences were demultiplexed and associated with each individual based on internal barcode sequences. SNP-calling was performed on samples from each species separately using DiscoSnp-RAD, a de novo reference-free and assembly-free method (Gauthier et al., 2017; Uricaru et al., 2015). SNPs are identified from particular arrangements in the De Bruijn graph built using a k-mer size of 31 and a minimal coverage of two for each allele (Gauthier et al., 2017; Uricaru et al., 2015). Individuals with more than 90% missing genotypes were excluded (Table S1) resulting in a final data set consisting of 105 samples for *I. salapia* and 142 for *O. onega*. SNPs scored in at least 80% of the samples (i.e., sites with <20% missing data) and with a minor allele frequency above 0.01 were retained using vcftools (Danecek et al., 2011) resulting in a data set of 17,779 SNPs for *I. salapia* and 15,894 SNPs for *O. onega*.

2.4 | Population structure analyses

Population genetic structure was investigated using a subset of SNPs, where only one SNP per GBS locus was considered so as to minimize the effects of linkage disequilibrium that would occur within loci, as often recommended (Falush, Stephens, & Pritchard, 2003). A total of 8,219 SNP for *I. salapia* and 5,133 SNP for *O. onega* were retained. To investigate genetic structure, we used principal component analysis implemented for genetic data in the adegenet R package (Jombart & Ahmed, 2011). We used Bayesian admixture analysis implemented in Structure (Pritchard, Stephens, & Donnelly, 2000) to estimate admixture proportions, that is, the proportion of each individual's genome inherited from each of *K* hypothetical source populations. We ran analyses with *K* from 1 to 6 with three independent Markov chains each, using 200,000 steps and including 10,000 burnin steps. We checked the results obtained in each run to verify convergence of the chains to a stable posterior distribution. The most likely number of clusters was identified using Evanno's method (Evanno, Regnaut, & Goudet, 2005) implemented in Structure Harvester (Earl & vonHoldt, 2012).

2.5 | Genome-wide introgression and estimates of differentiation

To investigate introgression among each population pair and to search for loci potentially associated with reproductive isolation, we used a genomic cline approach using bgc (Gompert & Buerkle, 2012). Loci acting as barriers to gene flow and linked regions should exhibit reduced introgression into the foreign genomic background. Locus-specific introgression is characterized by the probability ϕ of being inherited from a given parental population (here, the Amazonian lineage; the probability of being inherited from the Andean lineage is therefore $1 - \phi$). These probabilities are compared to the genome-wide average probability, which corresponds to the hybrid index. Introgression patterns can be summarized by two locus-specific genomic cline parameters: α, the genomic centre parameter, and β, the genomic cline rate parameter. The genomic cline centre parameter α specifies an increase (positive values of α) or decrease (negative values of α) in the probability of ancestry of the focal population (here, the Amazonian lineage). Positive or negative α values denote an asymmetry in the direction of introgression with hybrids having increased or decreased ancestry from one or the other ancestral lineage, respectively. The genomic cline rate parameter β specifies the cline steepness, with an increase (positive values of β) or decrease (negative values of β) in the rate of transition from low to high probability of ancestry (Gompert & Buerkle, 2011). Positive or negative β values are associated with a high or low level of gene flow, respectively. We estimated the posterior probability distribution of hybrid indices and cline parameters with bgc. Markov chain Monte Carlo (MCMC) of 50,000 steps including 10,000 burnin steps for *I. salapia* samples and 100,000 with 30,000 burnin steps for *O. onega* samples were used to reach mixing, and convergence was verified graphically by plotting log-likelihood distributions. For a given SNP, outlier introgression from the

genome-wide average was identified as credible when the 95% credible intervals of the cline parameters α and β excluded zero. These SNPs deviating from global pattern should reflect unusual patterns of evolution acting on these loci. We used the admixture model implemented in entropy (version 1.2) to estimate admixture proportions and inter-taxon ancestry (Gompert et al., 2014). This model explicitly estimates the proportion of each individuals' genome where the two allele copies are derived from different source populations (i.e., the proportion of the genome with intertaxon ancestry). entropy also incorporates uncertainty in genotypes due to limited sequence coverage and sequencing errors. We fit the model using Markov chain Monte Carlo (MCMC). We ran the MCMC algorithm three times with 15,000 iterations following a 5,000 iteration burnin, and with a thinning interval of five. We assumed the number of source populations (K) was two.

Genome-wide weighted and per-SNP genetic differentiation F_{ST} were estimated using Weir and Cockerham's method (Weir & Cockerham, 1984) implemented in vcftools (Danecek et al., 2011). To do so, only samples from parental lineages, that is, samples from the initial localities distant from the hybrid zone, were kept. To identify outlier SNPs with elevated genetic differentiation, the threshold was fixed to the 95th percentile of F_{ST} distribution obtained by random sampling with replacement of 100,000 values.

2.6 | Wing pattern analyses

Photographs of dorsal and ventral sides of detached wings of 90 and 94 of the genotyped specimens of *I. salapia* and *O. onega*, respectively, were taken with a Nikon D90 digital camera and a 105 mm lens on a white background with a piece of millimeter paper for scale. For each specimen, dorsal and ventral patterns of fore- and hindwings were quantified using Colour Pattern Modeling (CPM, (Le Poul et al., 2014) as follows: wings were first extracted from their background, resulting in eight images per specimen (two wings [forewing and hindwing] × 2 lateral sides [left/right] × 2 vertical sides [ventral/dorsal]); for each image, wing pattern was described by semi-automatically categorizing wings into a finite number of colours (yellow, black, orange and white for *I. salapia*; white, black and orange for *O. onega*). Damaged wings were discarded, and when left and right wings were available only one randomly chosen side was used in subsequent analyses. Homologous wings were then aligned according to both shape and pattern (Le Poul et al., 2014), and a binary principal components analysis based on one-hot encoding of colours (i.e., where each colour is encoded by a string of bits among which only one takes the value 1) was performed on the colour of homologous pixels shared by all wings. Principal component (PC) scores were used as a quantitative measure of colour pattern in subsequent analyses.

2.7 | Admixture association mapping

To identify SNPs associated with variation in wing patterns, we performed association mapping using the Genome-wide Efficient

Mixed Model Association tool (GEMMA) (Zhou & Stephens, 2012, 2014). We used the multivariate linear mixed model (mvLMM) to test marker associations with multiple phenotypes and to estimate genetic correlations among complex phenotypes. To do so, we retained all wing pattern PCs that explained more than 1% of the variation in each species as variables. This included 14 variables accounting for 57.1% of the variation in wing pattern for *I. salapia* and 18 variables explaining 54.9% of the variation for *O. onega*. Both the variation linked to sex and the confounding effect of population structure (Freedman et al., 2004; Price et al., 2006) were integrated into the models by implementing a relatedness matrix between individuals generated using GEMMA (Zhou & Stephens, 2012) and the first population structure PC (obtained using adegenet R package (Jombart & Ahmed, 2011)) as a covariate. Analyses were carried out using the option -l mm 1 to perform a Wald test evaluating the probability of the null hypothesis that the marker effect sizes for all phenotypes were zero. For the identification of SNPs significantly associated with wing pattern, the threshold was fixed to a *p*-value adjusted using (a) a classical Bonferroni correction, which divides the significance threshold by the number of multiple comparisons, that is, the number of molecular markers multiplied by the number of variables; and (b) using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995). We considered SNPs identified by both of these correction methods as significantly associated with wing pattern.

To test whether outlier SNPs by the three approaches, i.e., differentiation genome-scan, introgression pattern and admixture mapping, are distributed randomly or if there is enrichment in shared outlier SNPs, we used two methods: a Pearson's Chi-squared test to compare shared outlier to a random distribution and a bootstrap of 1,000 random samplings among SNPs to estimate confidence interval of such shared SNPs.

2.8 | Similarities with other Lepidoptera genomes

Loci containing SNPs identified as outliers in the genetic differentiation or the differential introgression approaches were screened by BLAST against all annotated butterfly reference genomes in LepBase v4 (Challis, Kumar, Dasmahapatra, Jiggins, & Blaxter, 2016) to investigate the gene content of homologous genomic regions. This was performed using the BLASTn tool available on the LepBase platform (Priyam et al., 2019). Best hits, their location in reference genomes and genes were then investigated manually.

3 | RESULTS

Our reduced complexity genotype by sequencing approach coupled with Illumina sequencing produced 77.8 million reads distributed relatively evenly between samples with a mean of 340,940 reads per individual (sd: 157,991) for *I. salapia* samples and a mean of 270,602 reads per individual (sd: 103,734) for *O. onega*. From this sequencing data, SNP calling and filtering steps resulted in final data sets of

17,779 SNPs from 6,972 loci for *I. salapia* samples and 15,894 SNPs from 4,524 loci for *O. onega* samples.

3.1 | Population genetic structure

For *I. salapia* we considered two parental lineages, an Amazonian lineage corresponding to subspecies *aquinia* sampled from two sites, Km-26 Yurimaguas-Tarapoto (1) and San Miguel de Achinamiza (2), and an Andean lineage corresponding to subspecies *derasa* and sampled from one site Puente Aguas Verdes (5). Between these sites a zone that is geographically and altitudinally intermediate and where those lineages are in contact was sampled at two sites (sites Km-42 Tarapoto-Yurimaguas (i3) and La Florida (i4), hybrid zone) (Figure 1a). The genetic structure, identified by multivariate analyses on genetic data (Figure 2a) as well as Structure (Figure 2b), highlight that the

samples probably segregate in three, best clustering, or two groups (Evanno's method: higher ΔK of 5,037.42 for K = 3 and 718.12 for K = 2). These results shows that the two parental lineages are distinct and show a low level of admixture. They have a weighted genome-wide differentiation F_{ST} of 0.177 between the two. In the intermediate populations, we did not find putative F1 hybrids (i.e., individuals with hybrid index close to 0.5 and high intertaxon ancestry). Rather, four individuals (subpopulation 3.1, Figure 2b) were genetically similar to individuals from the parental Amazonian *aquinia* lineage, and the 28 other samples had equal levels of admixture with the majority of their genetic content associated with the Andean *derasa* population (Figure 2b). Hybrid indices estimated for these individuals is also indicative of this (Figure 2c).

The distribution of *O. onega* hybrids in the study area is somewhat similar to that observed in *I. salapia*. The genetic structure, identified by multivariate analyses on genetic data (Figure 2a) as
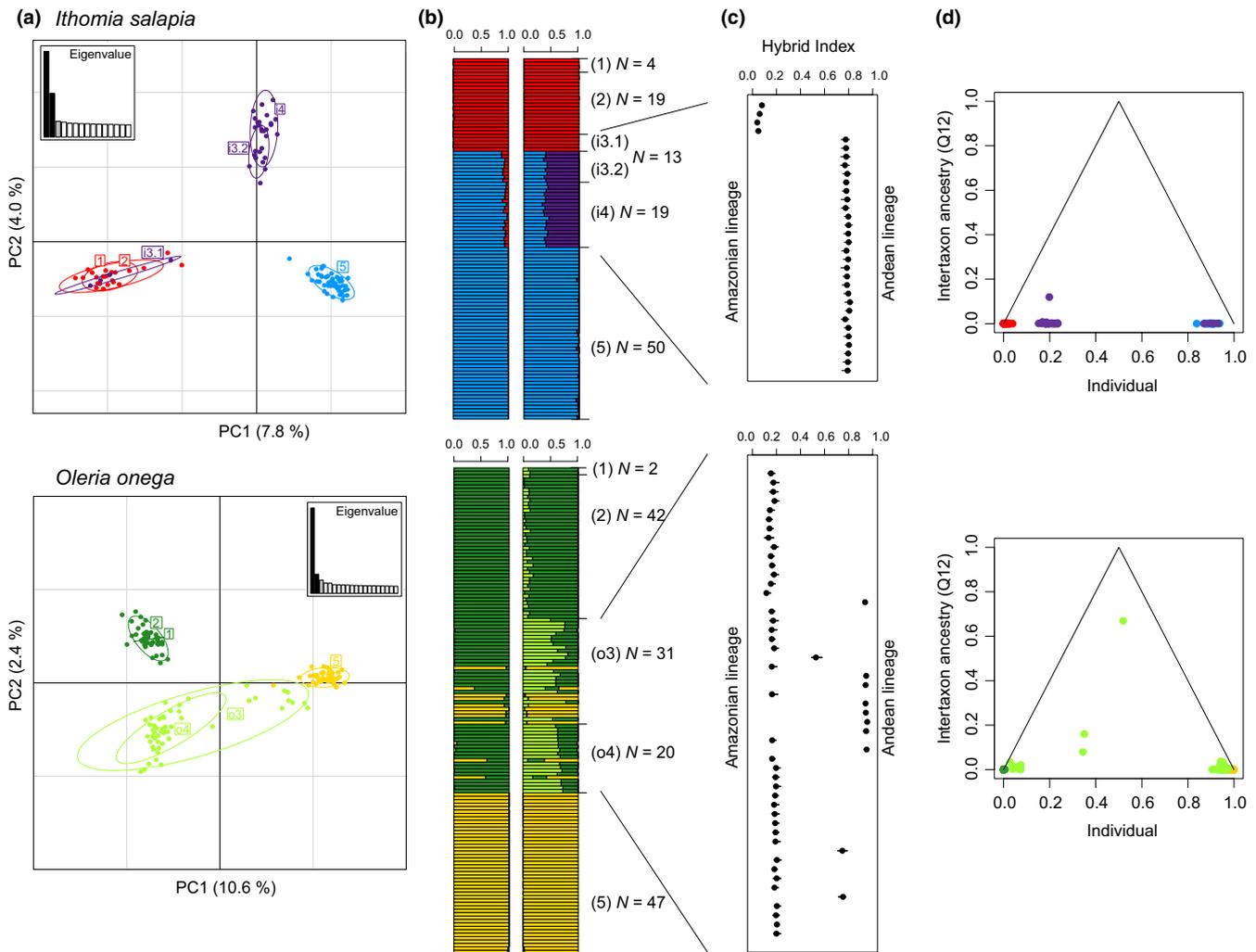

FIGURE 2 Population structure of pure and hybrid populations of *Ithomia salapia* (top) and *Oleria onega* (bottom). (a) Principal component analysis (PC1: horizontal axis, PC2: vertical axis), the percentage of total inertia explained by each axis is indicated in parentheses and the histograms in the top corners represent the inertia percentages of the first principal components. (b) Structure plot for K = 2 and K = 3. The number of individuals that were used is indicated for each site. (c) Plot of the hybrid index of each sample from the hybrid populations. The points represent the mean hybrid index value estimated from the posterior distribution and black lines indicate 95% credible intervals. (d) Plot of intertaxon ancestry and hybrid index. Population colour codes are the same as those in Figure 1b

well as Structure (Figure 2b), highlight the samples probably segregate in 2, best clustering, or three groups (Evanno's method: highest ΔK of 9,887.07 for K = 2 and 7.76 for K = 3). This species also consists of two parental lineages corresponding to an Amazonian subspecies, *O. onega janarilla*, collected from two sites, Km-26 Yurimaguas-Tarapoto (1) and San Miguel de Achinamiza (2) and an Andean subspecies, *O. onega* ssp. nov. 2 (Dasmahapatra et al., 2010; Gallusser, 2002), sampled at Puente Serranoyacu (5). In addition to these sites, a geographically and altitudinally intermediate zone was also sampled (sites Shapaja-Chazuta (o3) and a site spanning Quebrada Yanayacu to Laguna del Mundo Perdido (o4), hybrid zone) (Figure 1b). The two parental lineages are genetically divergent with a weighted genome-wide differentiation F_{ST} of 0.372, more than twice as high as between *I. salapia* parental lineages. The samples from the hybrid zone comprise a mix of individuals with low and intermediate levels of admixture. Most individuals have a low level of admixture (39 out of 46 had q < 0.2), and are genetically closest to the Amazonian parental lineage. The stronger link with the Amazonian lineage suggests a directionality in hybridization different from that of the *I. salapia* hybrid zone. Three individuals show hybrid indices that suggest almost equal contributions from each parental lineage (Figure 2a,b,c). One of these has a high level of intertaxon ancestry, which suggests it is an F1 hybrid. The two other individuals have lower heterozygosity, which is consistent with recent backcrossing. In conclusion, both species show some evidence of gene flow and introgression, but both also exhibit strongly bimodal phenotypes (sensu Jiggins & Mallet, 2000) in the region of the hybrid zone, suggesting strong reproductive isolation in both species. However, *O. onega* displays somewhat more evidence of ongoing hybridization and gene flow than *I. salapia*.

3.2 | Genomic patterns of introgression and differentiation

Introgression varied across the genome for each taxon pair, with distinct patterns of introgression for the two studied species, *I. salapia* and *O. onega*, as demonstrated by the distributions of the genomic cline centre (α) and rate parameters (β) (Figure 3). In each species the centre parameter (α) is highly variable, with point estimates (posterior median) ranging from −3.067 to 2.864 for *I. salapia* and from −5.785 to 5.907 for *O. onega*. With respect to α, many loci show introgression patterns that differ credibly from the genome-wide average. Loci with positive centre parameter α are more likely to be inherited from the Amazonian parental lineage than the rest of the genome. Conversely, loci with negative α are more likely to be inherited from the Andean lineage. For *I. salapia*, 2,125 SNPs (11.95% of all SNPs) have a centre parameter α that differ from the genome-wide average, consistent with different levels of introgression than the average of the rest of the genome. Among the SNPs with credible evidence of differential introgression, most of them (1,637 out of 2,125, i.e., 77%) have excess ancestry from the Amazonian populations whereas the genome-wide average is more closely associated

with the Andean population (Figure 2b,c). In *O. onega*, where intermediate populations are genetically closer to the Amazonian lineage, 2,146 SNPs (13.50% of all SNPs) have a centre parameter α that significantly deviates from the genome-wide average. The distribution is reversed compared to *I. salapia*, however, with most of those SNPs (1,406 out of 2,146, i.e., 66%) having a lower probability of being inherited from the Amazonian parental lineage. This is expected, as there is more statistical power to detect differential introgression of alleles of the less common (i.e., minor) ancestry type.

Regarding the genomic cline rate parameter (β), the profiles for the two species are markedly different. In *I. salapia* species, β hardly shows any variation, ranging from −0.681 to 0.738, whereas the variation observed in *O. onega* is higher by tenfold, ranging from −7.710 to 8.440. Such a large variation in *O. onega* is probably due to the heterogeneity of the hybridization profiles observed in the hybrid zone (Figure 2b). While in *I. salapia* no SNPs are different from the genome-wide expectation, in *O. onega*, 1,274 SNPs have a genomic cline rate (β) credibly different from the genome-wide pattern, with 447 and 827 SNPs having higher and lower values, respectively. These 447 SNPs with steeper introgression patterns than the genome-wide average are characteristic of SNPs putatively associated with barrier loci (i.e., in LD with barrier loci).

This difference observed between *I. salapia* and *O. onega* can be explained, at least in part, by the fact that there is hardly any variation among hybrids in *I. salapia* in our sampling. Perhaps an even larger sample than at present would show additional variation. In any case, the absence of variation in the *I. salapia* hybrids is a result in itself but limits the capacity to detect variation in patterns of introgression.

Genetic differentiation (F_{ST}) between parental lineages was heterogeneous across the genome, ranging from ~0 to 1.000 with a weighted mean value of 0.177 for *I. salapia* and from ~0 to 1.000 with a weighted mean value of 0.372 for *O. onega*. The genome scan also identifies outlier SNPs that deviate from the genome-wide distribution and have unusually high levels of differentiation. In *I. salapia*, the 95th percentile threshold corresponds to an F_{ST} value of 0.415, and was exceeded by 890 SNPs. In *O. onega*, the 95th percentile threshold corresponds to an F_{ST} value of 0.686, and was exceeded by 795 SNPs.

3.3 | Phenotypic variation

For both species, the first principal component on colour pattern separates Andean from Amazonian lineages (Figure 4). However, the two species differ in where along this axis specimens from intermediate populations fall. For *I. salapia*, most specimens in the hybrid zone cluster with the Andean *I. salapia derasa*, while only four (with a predominantly Amazonian genetic background) cluster with Amazonian *I. salapia aquinia* (Figure 4). No individual in this sample has a markedly intermediate colour pattern along this first principal component. By contrast, most *O. onega* from the hybrid zone have an intermediate position along the first principal component. For both
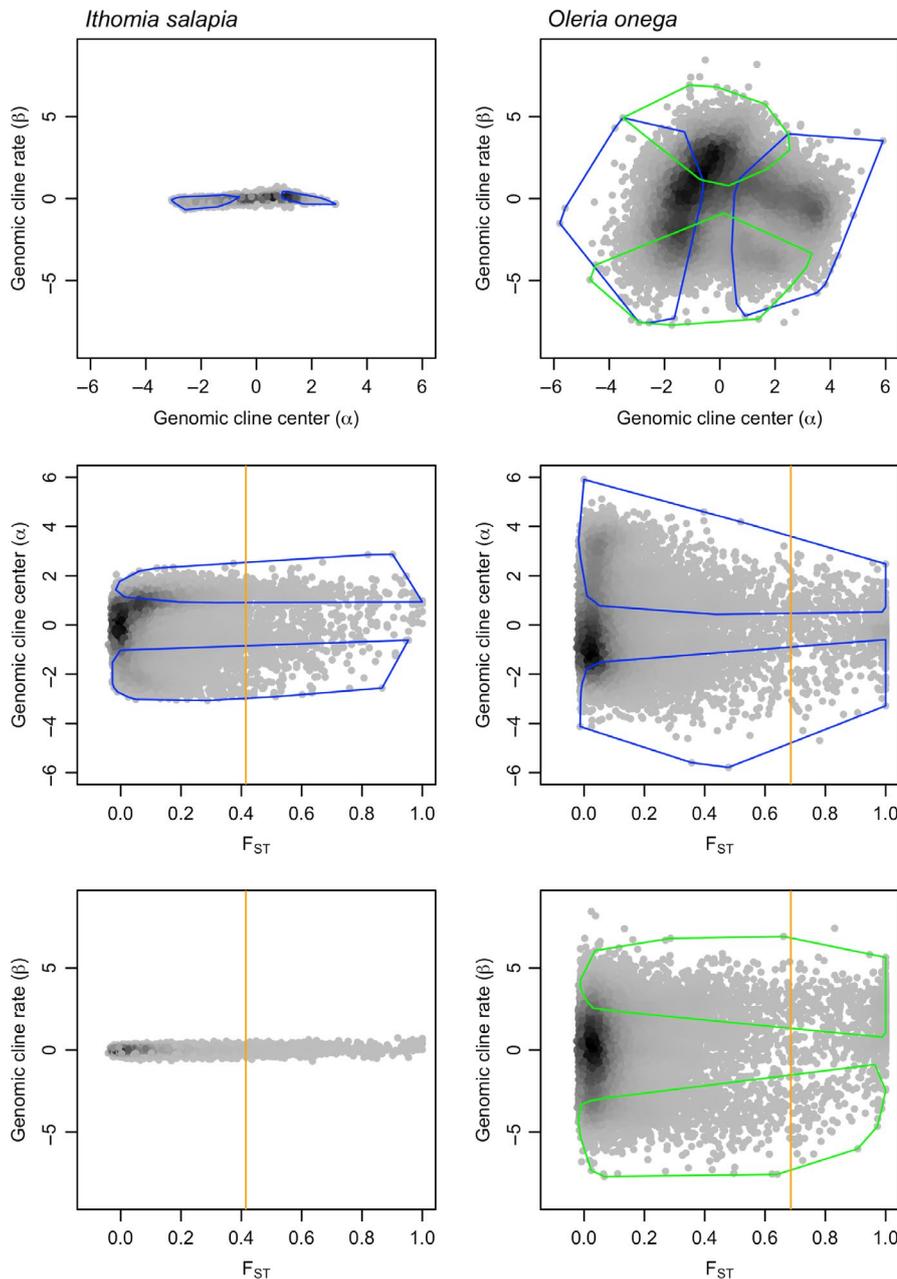
FIGURE 3 Scatterplots representing the relationships between the genomic cline center parameter (α), representing SNP ancestry; the genomic cline rate parameter (β), representing the steepness of the cline; and the differentiation level, F_{ST}, estimated for each SNP. Plots for *I. salapia* and *O. onega* are on theft and right, respectively. Each data point is coloured in grey, and darkness increases with point density (i.e., darker areas contain more points). Blue lines frame sets of SNPs for which the genomic cline center parameters (α) significantly deviates from the genome-wide pattern. Green lines frame sets of SNPs for which the genomic cline rate parameters (β) significantly deviates from the genome-wide pattern. Note that for *I. salapia*, no SNPs have genomic cline rate parameters (β) that deviate from the genome-wide pattern. SNPs on the right hand side of the orange lines harbour a significantly higher differentiation (high F_{ST}) than the genome average

species, the second axis highlights variation associated with sex and as a result males and females are segregated along this axis. While sexual dimorphism is moderate and of the same magnitude in both lineages of *O. onega*, it is more pronounced in the Andean lineage *I. salapia derasa* and virtually absent in the Amazonian lineage *I. salapia aquinia*. Sex was therefore included as a factor in our phenotype-genotype analyses, along with genetic structure.

3.4 | Phenotype-genotype relationship and association mapping

Despite the low variability among hybrid individuals from the hybrid zone, especially for *I. salapia*, we attempted to associate genetic variation in specific SNPs to wing pattern variation. All principal components (PCs) explaining at least 1% of total phenotypic variation were included in association mapping (i.e., 14 variables for *I. salapia*, jointly explaining 57.1% of the variance, and 18 for *O. onega*, jointly explaining 54.9% of the variance). Association mapping using GEMMA revealed several PCs for which a large phenotypic variation is explained, specifically more than 80% phenotypic variation explained (pve) for 7/14 PCs for *I. salapia* and for 4/18 PCs for *O. onega*. When combining all PCs that were retained (i.e., all those that explained at least 1% of the phenotypic variation), 59.3% of the phenotypic variation is explained for *I. salapia* and 26.3% for *O. onega* (Figure S2). The PCs explaining most genetic variation are also those harbouring the largest proportions of wing pattern variation. Specifically, 84.3% of the wing pattern variation is explained for *I. salapia* and 65.3% for *O. onega*. The multivariate linear mixed model of GEMMA performs tests to evaluate the probability that SNPs are associated
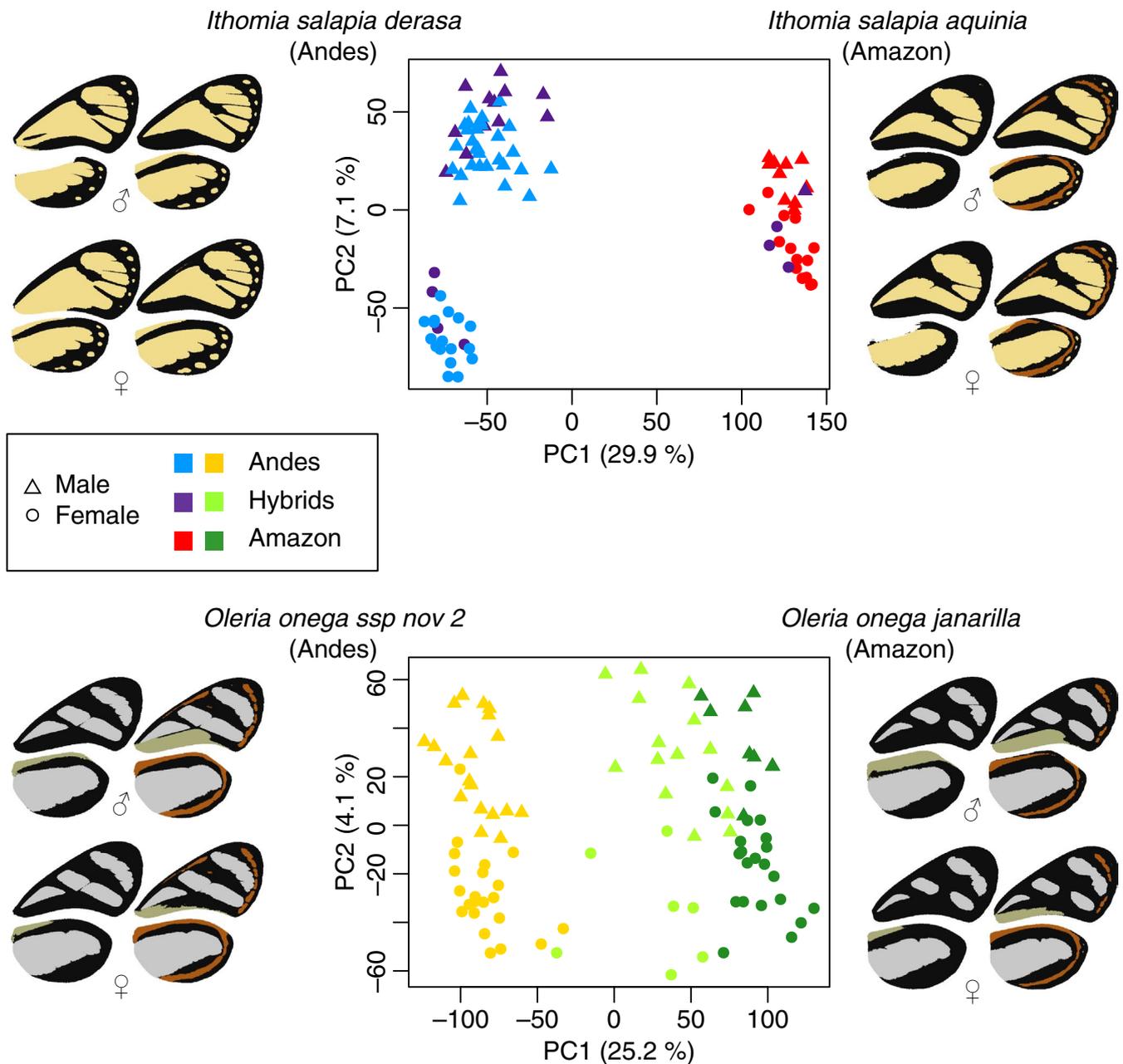
Ithomia salapia derasa (Andes)

Ithomia salapia aquinia (Amazon)

Oleria onega ssp nov 2 (Andes)

Oleria onega janarilla (Amazon)

△ Male
○ Female

Andes
Hybrids
Amazon

FIGURE 4 Phenotypic position of 90 *I. salapia* (top) and 94 *O. onega* (bottom) in the wing colour space consisting of the two main principal components from the colour pattern modelling approach. Colour indicates sample populations as in Figure 1b. Females and males are depicted by circles and triangles, respectively. Representative images of the average phenotypes for population and sex are shown on each side of the figure

with phenotypic variation and outputs the corresponding *p*-value resulting from a Wald test. Retaining significant SNPs concurrently in Bonferroni *p*-value correction and FDR approach, 88 SNPs (0.49% of all SNPs) were significantly associated with wing patterns in *I. salapia* and 109 SNPs (0.69% of all SNPs) in *O. onega* (Figure 5).

We then focused on the differentiation and introgression patterns of SNPs associated with wing pattern. A very small number of SNPs, 17 for *I. salapia* and 4 for *O. onega*, combine strong association with wing pattern and high levels of differentiation, but in *I. salapia* the SNP with the strongest association with wing pattern is also an outlier in genome scan for differentiation (Figure 5).

With all of these approaches combined, we observe both differences and similarities between *I. salapia* and *O. onega* in the number of SNPs similarly identified as outliers by multiple approaches. The proportion of SNPs combining both high differentiation levels (F_{ST}) and differential introgression (α), and which are characteristic of loci potentially involved in adaptation in hybrid samples, are of the same order in *I. salapia* and in *O. onega* (1.81%, 321/17,779, for *I. salapia*, and 1.59%, 253/15,894, for *O. onega*, Figure 6), and higher than expected at random (Pearson's Chi-squared test, *p* = 1.32e-114 and 95% CI = [103, 135] for *I. salapia* and Pearson's Chi-squared test, *p* = 2.99e-54 and 95% CI = [92, 123] for *O. onega*). SNPs previously identified as
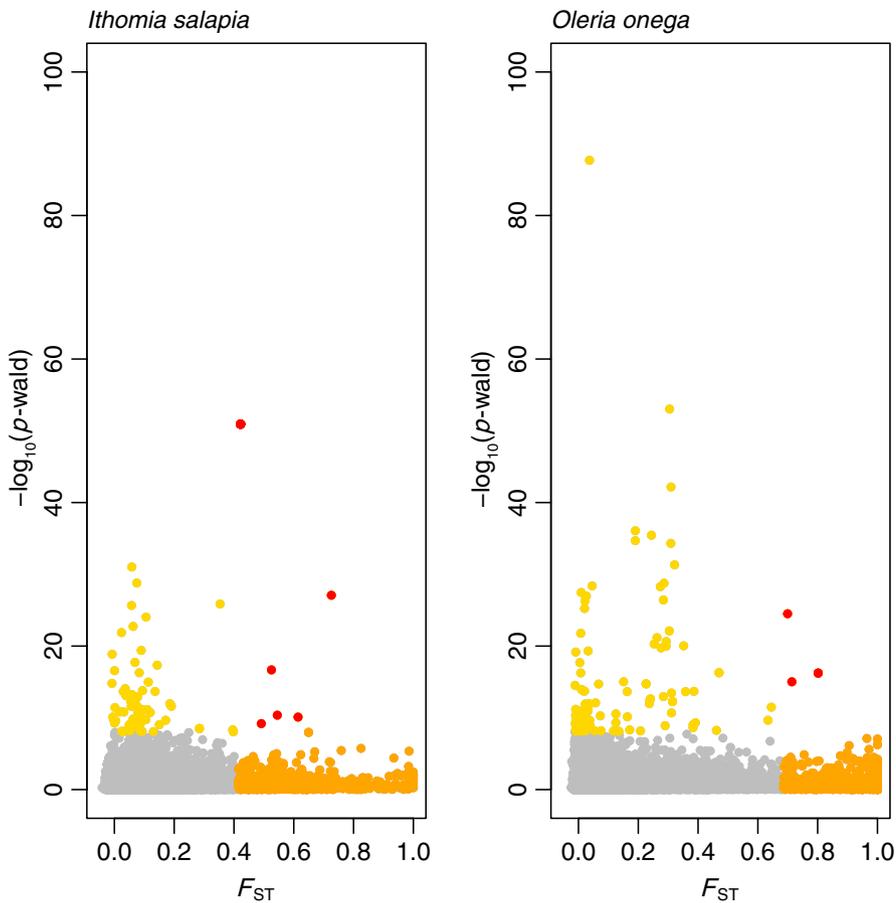
FIGURE 5 The relationship between the significance of association with colour pattern (represented as $-\log_{10}$(p-Wald)) and F_{ST} for *I. salapia* (left) and *O. onega* (right). Yellow points indicate SNPs significantly associated with wing patterns (after both Bonferroni and FDR corrections). Orange points highlight SNPs with high F_{ST} values and red points highlight SNPs with significant association both to wing pattern and to high F_{ST}
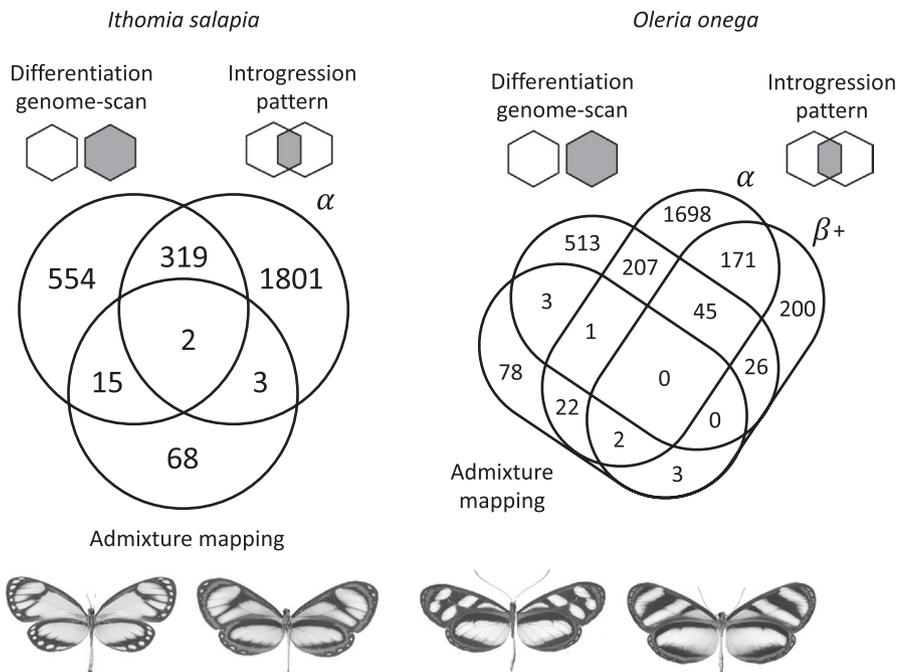

FIGURE 6 Venn diagram combining number of SNPs identified as supported by each approach (differentiation, introgression and admixture mapping) and shared between them. Note that for introgression patterns, no SNP showed deviant genomic cline rate parameters (β) in *I. salapia*. This parameter is therefore not represented in the diagram

potentially involved in adaptation and adaptive introgression, combining high differentiation levels (F_{ST}) and differential introgression (α), do not have a specific enrichment in SNPs significantly associated with wing pattern. In *I. salapia* and *O. onega* the numbers of SNPs that fit this description are low, respectively only two and one SNPs, and do not differ from a random distribution (Pearson's Chi-squared test, *p* = .74 and 95% CI = [0, 4] for *I. salapia* and Pearson's Chi-squared test, *p* = .57 and 95% CI = [0, 4] for *O. onega*). The main difference between the two species consists in the SNPs with differential positive genomic cline rate values (β) and potentially involved in reproductive isolation. None

of the SNPs in *I. salapia* have positive β while a nontrivial proportion do in *O. onega*, i.e., 447 SNPs or 2.81%. Among these SNPs only a small fraction is also significantly associated with wing pattern variation (5) and not enriched compared to a random distribution (Pearson's Chi-squared test, *p* = .26 and 95% CI = [1, 6]). Moreover, among the SNPs involved in both, high differentiation level and positive β (71) and potentially involved in reproductive isolation, none of them is associated with wing pattern variations (Figure 6).

3.5 | Similarities with other Lepidoptera genomes

We used BLASTn to investigate the potential functional roles of the loci carrying the SNPs highlighted by the approaches listed above. We identified homologous regions in the *D. plexippus* genome for loci containing SNPs significantly associated with wing patterns (12/67 loci for *I. salapia* and 21/99 for *O. onega*; Table S3). None of the genes known to control colour pattern variation in Lepidoptera and identified in the *D. plexippus* genome, i.e. *optix*, *cortex*, *WntA*, *ebony* and *aristaless*, were identified.

On the other side, the BLASTn of loci with SNPs potentially involved in adaptation and adaptive introgression or reproductive isolation highlighted candidate genomic region and gene in the genome of *D. plexippus* (37/277 loci for *I. salapia* and 66/218 for *O. onega* potentially involved in adaptation and adaptive introgression and 73/389 loci potentially involved in reproductive isolation for *O. onega*; Table S2). Here, we report the list of scaffold containing these loci of interest which are potential candidates for genes involved in local adaptation and reproductive isolation, and on which further functional analyses could be performed to investigate underlying biological functions (Table S2).

4 | DISCUSSION

The comparison of genome-wide patterns of genetic differentiation, introgression and genotype-phenotype associations in two species, *I. salapia* and *O. onega*, that face similar environmental transitions revealed some surprisingly large phenotypic and genomic differences. Below, we discuss potential reasons for the differences observed in light of biological and ecological information.

4.1 | Genomic and phenotypic differentiation patterns across the Tarapoto suture zone: Similarities and differences

Both *I. salapia* and *O. onega* are distributed across an important environmental gradient in the region of Tarapoto in Peru, and both species consist of an Andean and Amazonian lineage. Our analysis of wing pattern variation confirms that Amazonian and Andean lineages of both species are phenotypically different (as is seen by the human eye) and also reveals a subtle sexual dimorphism not readily discernible.

Phenotypic differentiation between populations of each species is associated with strong overall genomic differentiation, especially in *O. onega*. These findings are consistent with those obtained by Dasmahapatra et al. (2010) using four loci, which also revealed interlineage differentiation for these taxa, with the strongest genetic differentiation occurring in *O. onega*.

However, the genomic and phenotypic population structure of hybrid populations differ between *I. salapia* and *O. onega*. First, while all but four of the *I. salapia* individuals sampled in the hybrid zone are genetically closer to the Andean population, most individuals in the *O. onega* hybrid populations we sampled are genetically closer to Amazonian populations. Second, one individual in *O. onega* is probably a F1, and two other individuals are recent backcrosses, while no such genetically intermediate individuals were found in our samples of *I. salapia*. Third, the phenotypic structure of hybrid populations mirrors the genomic patterns. Along the first PC individuals in intermediate populations of *I. salapia* are phenotypically closest to the Andean parental lineage (*derasa*), to which they are also closest genetically. In our sample, intermediate colour patterns are not observed in these populations, nor in parental populations. By contrast, individuals in intermediate populations of *O. onega* have intermediate phenotypes between the two lineages, with a tendency to be closer to the Amazonian lineage (*janarilla*), to which they are also closest genetically.

Overall, the patterns detected suggest past gene flow in both species (most individuals have a similar, low hybrid index), with potentially more recent (but rare) gene flow in *O. onega* - although we cannot rule out the fact that we may have missed recent hybrids in *I. salapia*. Genomic differentiation across hybrid zones in Müllerian mimetic butterflies have mostly been documented in the genus *Heliconius*. While *Heliconius* subspecific lineages sometimes exhibit high genome-wide differentiation across hybrid zones (Martin, Davey, Salazar, & Jiggins, 2019; Van Belleghem et al., 2018), this appears not to be the case in the Tarapoto suture zone. In this region, Nadeau et al. (2014) found that in phenotypically differentiated lineages of *H. erato* and *H. melpomene* only loci around pattern gene loci showed genetic differentiation, while the rest of the genome was highly permeable to gene flow, with F_{ST} values ranging from 0.0112 to 0.0280 (see also Martin et al., 2013). This stands in stark contrast to the strong overall differentiation we revealed in ithomiine butterflies from the Tarapoto suture zone (F_{ST} = 0.177 for *I. salapia* and F_{ST} = 0.372 for *O. onega*).

While intermediate populations of *O. onega* show a high extent of genetic heterogeneity, in *I. salapia* all but four individuals from intermediate populations are remarkably similar in their genetic composition. This suggests that intermediate populations of *I. salapia* are hardly exchanging genes with Andean and Amazonian populations, and may be in the process of forming a distinct taxon.

4.2 | Genomic patterns of introgression

Variation in introgression patterns across the genome can help pinpoint loci involved in adaptation and reproductive isolation (Gompert & Buerkle, 2011; Gompert et al., 2012; Gompert, Mandeville, &

Buerkle, 2017). In particular, highly divergent SNPs with deviant genomic cline centre parameters (α) or positive genomic cline rate parameter (β) (i.e., exhibiting a steep cline) should be more common in regions of the genome involved in local adaptation or reproductive isolation.

Here, intermediate populations of both *I. salapia* and *O. onega* present SNPs with outlier values in their genomic cline centre parameters (α), meaning that these SNPs have an ancestry different from that of the average of the genome. The SNPs exhibiting deviant α should be enriched for genomic regions involved in adaptation or reproductive isolation. Such SNPs (77.0%) are shifted towards Amazonian ancestry in *I. salapia*, whereas the majority of SNPs with deviant α (65.5%) are shifted towards Andean ancestry in *O. onega*. While this may indicate introgression from the parental lineage that is least represented in the genomic background of intermediate populations, in our case such asymmetry may also result from a lower power to detect introgression from the dominant parental background. Whether some of those SNPs result from adaptive introgression, as has been revealed in *Heliconius* butterflies (Heliconius Genome Consortium, 2012; Jay et al., 2018), warrants further study.

Patterns of the parameter cline rate β markedly differ between *I. salapia* and *O. onega*. While in *I. salapia* no SNPs show outlier cline steepness, in *O. onega* many SNPs show narrower or wider clines compared to the genome average. Overall, in *O. onega*, the distribution of genomic cline parameters is wider and more heterogeneous than in *I. salapia*, suggesting less constraints in the hybridization process. Such heterogeneity in *O. onega* allows the identification of loci with specific introgression levels. Highly divergent genomic regions that have low levels of introgression are probably associated with reproductive isolation (Gompert & Buerkle, 2011; Gompert et al., 2012). Low levels of introgression can be the result of several evolutionary processes involving both extrinsic mechanisms, such as divergent selection and environment-dependent selection against hybrids, and intrinsic mechanisms such as an environment-independent reduced hybrid fitness caused by Bateson-Dobzhansky-Muller incompatibilities (Gompert & Buerkle, 2011; Gompert et al., 2012). Correlation of genetic patterns with other evidence (e.g., candidate traits) may shed light on the mechanisms of speciation and reproductive isolation (Ravinet et al., 2017).

4.3 | Genetic bases of colour pattern variation

Our admixture mapping analysis revealed SNPs associated with colour pattern in *I. salapia* (88 SNPs, representing 0.49% of all SNPs) and *O. onega* (109 SNPs, representing less than 0.69% of all SNPs).

In nymphalid butterflies, wing pattern variation can be explained by combinations of conserved pattern elements (Martin & Reed, 2014) and tends to be controlled by small numbers of loci (Van Belleghem et al., 2017; Zhang et al., 2017). Previous studies, including studies on mimetic *Heliconius* (Joron et al., 2006; Martin et al., 2012; Nadeau, 2016; Reed et al., 2011; Westerman et al., 2018) and

Papilio (Timmermans et al., 2014) identified a list of candidate genes such as *WntA*, *optix*, *cortex*, *ebony* and *aristaless*. Some of these have been functionally characterized (Martin & Reed, 2014; Nadeau, 2016; Nadeau et al., 2016). A recent study on *D. plexippus*, the most closely related species to Ithomiini for which a reference genome is available, highlighted the role of *WntA* in vein shape (Mazo-Vargas et al., 2017).

None of our candidate loci correspond to genes known to be involved in wing colour pattern in other butterflies. This is probably due to the relatively low-resolution genotype-by-sequencing approach adopted here, such that we may have missed gene regions that were not covered by our loci.

Moreover, only a small fraction of loci with SNPs associated with wing pattern (23.9% for *I. salapia* and 21.2% for *O. onega*) map to an orthologous region in the *D. plexippus* genome. This deficit is related to the relatively large divergence time between our focal species and *D. plexippus* (ca. 42 million years ago, Chazot et al., 2019), which limits our ability to find orthologous regions and more specifically to find regions involved in noncoding regulatory loci. We may therefore have missed loci that contain known genes involved in wing pattern development.

Finally, the extremely low level of hybridization observed in *I. salapia* reduces the statistical power of admixture mapping and hampers detection of genomic regions associated with wing pattern variation. The function of most regions identified in our analyses are unknown and represent a starting point for further analyses of these regions, as those regions may contain novel genes in these pathways.

4.4 | Colour pattern and reproductive isolation

Wing colour pattern is known to cause pre- and post-zygotic reproductive isolation in Müllerian mimetic butterflies (e.g., Chamberlain, Hill, Kapan, Gilbert, & Kronforst, 2009; Jiggins et al., 2001; Mallet & Barton, 1989; Merrill et al., 2011; Merrill et al., 2012; Naisbit, Jiggins, Linares, Salazar, & Mallet, 2002), including Ithomiini (McClure et al., 2019).

In our admixture mapping analysis, we found that only two and one of the significantly differentiated introgression outliers were also associated with wing pattern variation in *I. salapia* and *O. onega*, respectively. These figures do not differ from random expectations. These results suggest that wing colour pattern may be moderately involved in reproductive isolation in both species, but since our genomic data do not cover the entire genome, we cannot rule out the fact that we may have missed some important loci involved in wing pattern coloration and with deviant genomic clines.

In mimetic butterflies, hybrid individuals with intermediate colour pattern may suffer more predation because they are not recognized as unpalatable (e.g., Merrill et al., 2012), which may in turn select for assortative mating for wing colour pattern through reinforcement (e.g., Kronforst, Young, & Gilbert, 2007), resulting in reproductive isolation between phenotypically differentiated lineages.

Whether individuals with intermediate phenotype suffer increased predation has never been tested in *I. salapia* and *O. onega*, but predation experiments on *Heliconius* species carried out in the same region demonstrated the ability of predators to discriminate fine phenotypic differences (Arias et al., 2016; Chouteau, Arias, & Joron, 2016). Assortative mating seems likely in *I. salapia* and *O. onega* (M. McClure and M. Elias, personal observation), and has been documented by genetic and phenotypic characterization of the reared offspring of females collected in hybrid populations of *O. onega* (De Silva, 2010: chapter 5).

There are fewer phenotypically intermediate individuals in *I. salapia* than in *O. onega*. This difference might be explained by the mimicry rings to which the two species belong. While the mimicry rings of *I. salapia* lineages are readily discriminated and show little variation within each mimicry ring, the forms *O. onega* belongs to are more variable with overlapping phenotypes (M. Elias personal observation; Figure S1). Because of the greater variation and overlap of the two *O. onega* mimicry rings in Tarapoto, selection against hybrids with intermediate phenotypes may be reduced compared with that in *I. salapia*, thereby allowing the persistence of greater levels of gene flow between lineages. Whether the absence of intermediate phenotypes in *I. salapia* is due to high mortality of hybrids through predation, strong assortative mating, hybrid incompatibilities or all of these is currently unknown and deserves further examination.

4.5 | Other putative adaptive traits

In both species, only a small number of SNPs potentially involved in adaptation or reproductive isolation (i.e., highly differentiated SNPs that show deviant α or significantly positive β) are also associated with wing pattern. This suggests that other traits may play a role in reproductive isolation. Because interactions with local host plants at the larval stage and the ability to fully exploit them often impact fitness in phytophagous insects (Simon et al., 2015), larval host plant shifts are believed to be an important driver of reproductive isolation. However, in *O. onega* the two lineages utilize the same larval hostplants, *Solanum mite* and related *Solanum* sect. *Pteroidea* species (de-Silva, Vásquez, & Mallet, 2011; Gallusser, 2002). Similarly, *I. salapia derasa* larvae commonly feed on *Witheringia solanacea* (Beccaloni, 1997), a plant also used by *I. salapia aquinia* (JM, MM and ME, unpublished observations). Shifts in hostplant are therefore unlikely to explain divergence between lineages in either of these species, as is the case in another ithomiine genus, *Melinaea*, present in the same region (McClure & Elias, 2016).

The two lineages of *O. onega* have divergent egg-laying behaviour: females of the Amazonian population (*janarilla*) lay eggs on the hostplants, while females of the Andean population (ssp. nov. 2) tend to lay eggs off the host plant (Gallusser, 2002). Eggs are typically laid up to 0.5 m away from the nearest host plant individual, on twigs, leaf litter or live nonhost plant, which reduces egg predation (de-Silva et al., 2011). Differences in egg-laying behaviour have been shown to cause reduced hybrid fitness in butterflies (McBride

& Singer, 2010). This could be the case here, too, if hybrid females lay eggs off the plant, and if first instar larvae are incapable of locating their host plant.

Other putative adaptive traits include adaptations to distinct habitats (higher elevations and cooler temperatures for Andean lineages) and potentially microhabitats where comimics are most abundant (e.g., Elias, Gompert, Jiggins, & Willmott, 2008).

Finally, as many butterfly species, Ithomiini probably rely on sexual pheromones during mate choice (Schulz et al., 2004), and differences in sexual pheromones may incur discrimination between lineages. Notably, putative male pheromones have been shown to differ between the two lineages of *O. onega* (Stamm, Mann, McClure, Elias, & Schulz, 2019).

The role of these traits in reproductive isolation remains to be further explored using both experimental and genomic approaches.

AUTHOR CONTRIBUTIONS

M.E., and Z.G. designed the study. L.dS., J.M., M.M., and M.E. performed sampling. L.dS., A.W., Z.G., and M.E. performed labwork. J.G. analysed the molecular data, with contributions from Z.G., A.W., C.L., and F.L. M.E., J.G., C.H., and Y.L.P. analysed phenotypic data. All authors took part in discussions concerning the analyses and result interpretations. J.G. wrote the paper, with contributions from all authors.

DATA AVAILABILITY STATEMENT

Sequence reads are archived at the NCBI SRA in the BioProject PRJNA575968. Scripts describing the whole analytic process have been uploaded to GitHub (https://github.com/JeremyLGauthier/Scripts_Gauthier_et.al_2019_ME).

ORCID

Jérémy Gauthier https://orcid.org/0000-0001-6666-1002
Zachariah Gompert https://orcid.org/0000-0003-2248-2488

Annabel Whibley https://orcid.org/0000-0003-1878-7705
Yann Le Poul https://orcid.org/0000-0002-5013-255X
Melanie McClure https://orcid.org/0000-0003-3590-4002
Claire Lemaitre https://orcid.org/0000-0001-8675-170X
Fabrice Legeai https://orcid.org/0000-0002-6472-4839
James Mallet https://orcid.org/0000-0002-3370-0367
Marianne Elias https://orcid.org/0000-0002-1250-2353

REFERENCES

Arias, M., le Poul, Y., Chouteau, M., Boisseau, R., Rosser, N., Théry, M., & Llaurens, V. (2016). Crossing fitness valleys: Empirical estimation of a fitness landscape associated with polymorphic mimicry. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829), 20160391.

Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(3), 357.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497–503.

Bates, H. W. (1862). XXXII. Contributions to an insect fauna of the Amazon valley. *Lepidoptera: Heliconidae. Transactions of the Linnean Society of London*, *23*(3), 495–566.

Beccaloni, G. W. (1997). Ecology, natural history and behaviour of Ithomiine butterflies and their mimics in Ecuador (Lepidoptera: Nymphalidae: Ithomiinae). *Tropical Lepidoptera Research*, *8*(2), 103–124.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289–300.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*(10), 2044–2072.

Brown, K. S., Sheppard, P. M., & Turner, J. R. G. (1974). Quaternary refugia in tropical America: Evidence from race formation in *Heliconius* butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *187*(1088), 369–378.

Buerkle, C. A., & Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution*, *23*(12), 686–694.

Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., & Blaxter, M. (2016). *Lepbase: The Lepidopteran genome database* (p. 056994). https://doi.org/10.1101/056994

Chamberlain, N. L., Hill, R. I., Kapan, D. D., Gilbert, L. E., & Kronforst, M. R. (2009). Polymorphic butterfly reveals the missing link in ecological speciation. *Science*, *326*(5954), 847–850.

Chazot, N., Wahlberg, N., Freitas, A. V. L., Mitter, C., Labandeira, C., Sohn, J.-C., … Heikkilä, M. (2019). Priors and posteriors in Bayesian timing of divergence analyses: The age of butterflies revisited. *Systematic Biology*, *68*(5), 797–813.

Chouteau, M., Arias, M., & Joron, M. (2016). Warning signals are under positive frequency-dependent selection in nature. *Proceedings of the National Academy of Sciences*, *113*(8), 2164–2169.

Christe, C., Stölting, K. N., Bresadola, L., Fussi, B., Heinze, B., Wegmann, D., & Lexer, C. (2016). Selection against recombinant hybrids maintains reproductive isolation in hybridizing Populus species despite F1 fertility and recurrent gene flow. *Molecular Ecology*, *25*(11), 2482–2498.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., & DePristo, M. A. … 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Dasmahapatra, K. K., Lamas, G., Simpson, F., & Mallet, J. (2010). The anatomy of a "suture zone" in Amazonian butterflies: A coalescent-based test for vicariant geographic divergence and speciation. *Molecular Ecology*, *19*(19), 4283–4301.

de-Silva, D. L. (2010). *Ecology and Evolution in Neotropical Butterflies of the Subtribe Oleriina (Lepidoptera: Nymphalidae: Danainae: Ithomiini)*. PhD thesis, University of London, 275 pp.

De-silva, D. L., Vásquez, A. S., & Mallet, J. (2011). Selection for enemy-free space: Eggs placed away from the host plant increase survival of a neotropical ithomiine butterfly. *Ecological Entomology*, *36*(6), 667–672.

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.

Elias, M., Gompert, Z., Jiggins, C., & Willmott, K. (2008). Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. *PLOS Biology*, *6*(12), 2642–2649.

Endler, J. A. (1977). Geographic variation, speciation, and clines. *Monographs in Population Biology*, *10*, 1–246.

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.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., … Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, *36*(4), 388–393.

Gallusser, S. A. (2002). *Biology, behaviour and taxonomy of two* Oleria onega *subspecies (Ithomiinae, Nymphalidae, Lepidoptera) in north-eastern, Peru (Université de Neuchâtel)*. Retrieved from http://doc.rero.ch/record/2627

Gauthier, J., Mouden, C., Suchan, T., Alvarez, N., Arrigo, N., Riou, C., … Peterlongo, P. (2017). *DiscoSnp-RAD: De novo detection of small variants for population genomics* (p. 216747). https://doi.org/10.1101/216747

Gompert, Z., & Buerkle, C. A. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, *18*(6), 1207–1224.

Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, *20*(10), 2111–2127.

Gompert, Z., & Buerkle, C. A. (2012). bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, *12*(6), 1168–1176.

Gompert, Z., & Buerkle, C. A. (2013). Analyses of genetic ancestry enable key insights for molecular ecology. *Molecular Ecology*, *22*(21), 5278–5294.

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*(18), 4555–4573.

Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, *66*(7), 2167–2181.

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207–229.

Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98.

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., … Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, *28*(11), 1839–1845.e3.

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, *15*, 250–255.

Jiggins, C. D., Naisbit, R. E., Coe, R. L., & Mallet, J. (2001). Reproductive isolation caused by colour pattern mimicry. *Nature*, *411*(6835), 302–305.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61.

Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., ... Jiggins, C. D. (2006). A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLOS Biology*, *4*(10), e303.

Kapan, D. D. (2001). Three-butterfly system provides a field test of müllerian mimicry. *Nature*, *409*(6818), 338–340.

Kozak, K. M., Wahlberg, N., Neild, A. F. E., Dasmahapatra, K. K., Mallet, J., & Jiggins, C. D. (2015). Multilocus species trees show the recent adaptive radiation of the mimetic H*eliconius* butterflies. *Systematic Biology*, *64*(3), 505–524.

Kronforst, M. R., Hansen, M. E. B., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., ... Mullen, S. P. (2013). Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports*, *5*(3), 666–677.

Kronforst, M. R., Young, L. G., & Gilbert, L. E. (2007). Reinforcement of mate preference among hybridizing *Heliconius* butterflies. *Journal of Evolutionary Biology*, *20*(1), 278–285.

Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution*, *67*(12), 3653–3661.

Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., & Joron, M. (2014). Evolution dominance mechanisms at a butterfly mimicry supergene. *Nature Communications*, *5*, 5644.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237.

Mallet, J., & Barton, N. H. (1989). Strong natural selection in a warning-color hybrid zone. *Evolution*, *43*(2), 421–431.

Martin, A., Papa, R., Nadeau, N. J., Hill, R. I., Counterman, B. A., Halder, G., ... Reed, R. D. (2012). Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences*, *109*(31), 12632–12637.

Martin, A., & Reed, R. D. (2014). Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Developmental Biology*, *395*(2), 367–378.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*, 1817–1828.

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology*, *17*(2), e2006288.

Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R. W. R., Zhang, L., ... Martin, A. (2017). Macroevolutionary shifts of WntA function potentiate butterfly wing-pattern diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(40), 10701–10706.

McBride, C. S., & Singer, M. C. (2010). Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLOS Biology*, *8*(10), e1000529.

McClure, M., & Elias, M. (2016). Unravelling the role of host plant expansion in the diversification of a Neotropical butterfly genus. *BMC Evolutionary Biology*, *16*(1), 128.

McClure, M., Mahrouche, L., Houssin, C., Monllor, M., Le Poul, Y., Frérot, B., ... Elias, M. (2019). Does divergent selection predict the evolution of mate preference and reproductive isolation in the tropical butterfly genus *Melinaea* (Nymphalidae: Ithomiini)? *The Journal of Animal Ecology*, *88*, 940–952. https://doi.org/10.1111/1365-2656.12975

Merrill, R. M., Gompert, Z., Dembeck, L. M., Kronforst, M. R., McMillan, W. O., & Jiggins, C. D. (2011). Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution; International Journal of Organic Evolution*, *65*(5), 1489–1500.

Merrill, R. M., Wallbank, R. W. R., Bull, V., Salazar, P. C. A., Mallet, J., Stevens, M., & Jiggins, C. D. (2012). Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4907–4913.

Moritz, C., Hoskin, C. J., MacKenzie, J. B., Phillips, B. L., Tonione, M., Silva, N., ... Graham, C. H. (2009). Identification and dynamics of a cryptic suture zone in tropical rainforest. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1660), 1235–1244.

Müller, F. (1879). *Ituna* and *Thyridia*; a remarkable case of mimicry in butterflies. *Transactions of the Entomological Society of London*, *1879*, 20-29.

Nadeau, N. J. (2016). Genes controlling mimetic colour pattern variation in butterflies. *Current Opinion in Insect Science*, *17*, 24–31.

Nadeau, N. J., Pardo-Diaz, C., Whibley, A., Supple, M. A., Saenko, S. V., Wallbank, R. W. R., ... Jiggins, C. D. (2016). The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature*, *534*(7605), 106–110.

Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., ... Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research*, *24*(8), 1316–1333.

Naisbit, R. E., Jiggins, C. D., Linares, M., Salazar, C., & Mallet, J. (2002). Hybrid sterility, Haldane's rule and speciation in *Heliconius cydno* and *H. melpomene*. *Genetics*, *161*(4), 1517–1526.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402.

Pallares, L. F., Harr, B., Turner, L. M., & Tautz, D. (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. *Molecular Ecology*, *23*(23), 5756–5770.

Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21*(12), 2991–3005.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Priyam, A., Woodcroft, B. J., Rai, V., Munagala, A., Moghul, I., Ter, F., ... Wurm, Y. (2019). Sequenceserver: A modern graphical user interface for custom BLAST databases. *Molecular Biology and Evolution*, *36*(12), 2922–2924.

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., ... McMillan, W. O. (2011). *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, *333*(6046), 1137–1141.

Remington, C. L. (1968). Suture-zones of hybrid interaction between recently joined biotas. In T. Dobzhansky, M. K. Hecht, & W. C. Steere (Eds.), *Evolutionary Biology* (pp. 321–428). Boston, MA: Springer.

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*(2), 713–727.

Rissler, L. J., & Smith, W. H. (2010). Mapping amphibian contact zones and phylogeographical break hotspots across the United States. *Molecular Ecology*, *19*(24), 5404–5416.

Roberts, J. L., Brown, J. L., von May, R., Arizabal, W., Schulte, R., & Summers, K. (2006). Genetic divergence and speciation in lowland and montane peruvian poison frogs. *Molecular Phylogenetics and Evolution*, 41(1), 149–164.

Safran, R. J., & Nosil, P. (2012). Speciation: The origin of new species. *Nature Education Knowledge*, 3(10), 17.

Schulz, S., Beccaloni, G., Brown, K. S. Jr, Boppre, M., Freitas, A. V. L., Ockenfels, P., & Trigo, J. R. (2004). Semiochemicals derived from pyrrolizidine alkaloids in male ithomiine butterflies (Lepidoptera: Nymphalidae: Ithomiinae). *Biochemical Systematics and Ecology*, 32(8), 699–713.

Simon, J.-C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquiéry, J., Nouhaud, P., ... Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, 14(6), 413–423.

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., ... Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, 515(7527), 406–409.

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738–742.

Stamm, P., Mann, F., McClure, M., Elias, M., & Schulz, S. (2019). Chemistry of the androconial secretion of the Ithomiine butterfly *Oleria onega*. *Journal of Chemical Ecology*, 45(9), 768–778.

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., ... Tucker, P. K. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, 18(1), 67–76.

Teeter, K. C., Thibodeau, L. M., Gompert, Z., Buerkle, C. A., Nachman, M. W., & Tucker, P. K. (2010). The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution; International Journal of Organic Evolution*, 64(2), 472–485.

Timmermans, M. J. T. N., Baxter, S. W., Clark, R., Heckel, D. G., Vogel, H., Collins, S., ... Vogler, A. P. (2014). Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787), 20140465.

Uricaru, R., Rizk, G., Lacroix, V., Quillery, E., Plantard, O., Chikhi, R., ... Peterlongo, P. (2015). Reference-free detection of isolated SNPs. *Nucleic Acids Research*, 43(2), e11.

Van Belleghem, S. M., Baquero, M., Papa, R., Salazar, C., McMillan, W. O., Counterman, B. A., ... Martin, S. H. (2018). Patterns of Z chromosome divergence among *Heliconius* species highlight the importance of historical demography. *Molecular Ecology*, 27(19), 3852–3872.

Van Belleghem, S. M., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., ... Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, 1(3), 52.

Via, S., & Hawthorne, D. J. (2002). The genetic architecture of ecological specialization: Correlated gene effects on host use and habitat choice in pea aphids. *The American Naturalist*, 159(Suppl 3), S76–S88.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370.

Weir, J. T. (2006). Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. *Evolution*, 60(4), 842–855.

Westerman, E. L., VanKuren, N. W., Massardo, D., Tenger-Trolander, A., Zhang, W., Hill, R. I., ... Kronforst, M. R. (2018). Aristaless controls butterfly wing color variation used in mimicry and mate choice. *Current Biology*, 28(21), 3469–3474.e4.

Whinnett, A., Zimmermann, M., Willmott, K. R., Herrera, N., Mallarino, R., Simpson, F., ... Mallet, J. (2005). Strikingly variable divergence times inferred across an Amazonian butterfly "suture zone". *Proceedings of the Royal Society B: Biological Sciences*, 272(1580), 2525–2533.

Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A., & Reed, R. D. (2017). Genetic basis of melanin pigmentation in butterfly wings. *Genetics*, 205(4), 1537–1550.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821–824.

Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, 11(4), 407–409.

SUPPORTING INFORMATION

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

How to cite this article: Gauthier J, de Silva DL, Gompert Z, et al. Contrasting genomic and phenotypic outcomes of hybridization between pairs of mimetic butterfly taxa across a suture zone. *Mol Ecol*. 2020;00:1–16. https://doi.org/10.1111/mec.15403