

Genetic, morphological and ecological variation across a sharp hybrid zone between two alpine butterfly species

Thibaut Capblancq^{1,2}  | Laurence Després¹  | Jesús Mavárez^{1,3}

¹Laboratoire d'Écologie Alpine, UMR UGA-USMB-CNRS 5553, Université Grenoble Alpes, Grenoble, France

²Department of Plant Biology, University of Vermont, Burlington, VT, USA

³Departamento de Ciencias Biológicas y Ambientales, Universidad Jorge Tadeo Lozano, Bogotá, Colombia

Correspondence

Thibaut Capblancq, Laboratoire d'Écologie Alpine, UMR UGA-USMB-CNRS 5553, Université Grenoble Alpes, Grenoble 38000, France.

Email: thibaut.capblancq@gmail.com

Funding information

CNRS INEE, Grant/Award Number: APEGE; Conseil Général de l'Isère

Abstract

Identifying the mechanisms involved in the formation and maintenance of species is a central question in evolutionary biology, and distinguishing the selective drivers of populations' divergence from demographic processes is of particular interest to better understand the speciation process. Hybrid zones are recognized to provide ideal places to investigate the genetic architecture of speciation and to identify the mechanisms allowing diverging species to maintain their integrity in the face of gene flow. Here, we studied two alpine butterfly species, *Coenonympha macromma* and *C. gardetta*, which can be found flying together and hybridizing in narrow contact zones in the southern French Alps. We characterized the genomic composition of individuals, their morphology and their local habitat requirements, within and around a hybrid zone. Genetic diversity analysis at 794 SNPs revealed that all individuals within the hybrid zone were highly admixed, which was not the case outside the hybrid zone. Cline analysis showed that, despite ongoing hybridization, 56 out of 122 loci differentially fixed or nearly so between the two species were impermeable to introgression across the sharp hybrid zone (9 km wide). We also found concordance in cline position and width among genetic, morphological and environmental variation, suggesting a coupling of different reproductive barriers. Habitat characteristics such as the presence of trees and shrubs and the start of the growing season were strongly associated with the genetic variation, and we found evidence of divergence at genetic markers associated with morphology and physiology, putatively involved in visual or environmental reproductive isolation. We discuss the various behavioural and ecological factors that might interplay to maintain current levels of divergence and gene flow between this species pair.

KEYWORDS

Alps, *Coenonympha*, genetic cline, hybrid zone, morphometrics, speciation

1 | INTRODUCTION

Hybrid zones are geographic regions where genetically divergent taxa meet and hybridize (Barton & Hewitt, 1985). They are maintained

by a balance between divergent selection, promoting species differentiation, and gene flow, acting as a homogenizing force between genetic pools (Barton, 2001; Barton & Hewitt, 1985; Endler, 1977; Haldane, 1948; Jiggins & Mallet, 2000). In such zones, genes can

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

potentially be exchanged between populations through segregation and recombination in hybrids, at rates that, depending on the intensity of selection, can go from virtually zero for the genomic regions highly involved in species isolation and differentiation, to almost free circulation for the “neutral” or not divergently selected parts of the genome (Barton, 1979; Harrison, 1986, 1993; Payseur, 2010; Rieseberg, Whitton, & Gardner, 1999). Hybrid zones thus provide natural segregation experiments that allow both the study of factors that shape gene flow between divergent lineages and the identification of genetic loci and traits that contribute to species isolation (Payseur, 2010). The analysis of hybrid zones has therefore played an important role in the understanding of the genomic regions under selection and/or responsible for isolation in species that are difficult to breed in the laboratory (e.g. insects with a diapause) or long-lived organisms (e.g. trees) (Derryberry, Derryberry, Maley, & Brumfield, 2014; Lindtke et al., 2012).

The estimation of cline width and slope for a genetic locus along the hybrid zone provides important information about its introgression rate. When performed over many loci distributed throughout the genome, this analysis not only gives genome-wide introgression rates, but also allows for the identification of genomic regions exhibiting significantly higher-than-average or lower-than-average introgression between species (Abbott et al., 2013; De La Torre, Ingvarsson, & Aitken, 2015; Gay, Crochet, Bell, & Lenormand, 2008). The study of such patterns of heterogeneity in genetic exchange along the genome can help identify barrier loci, and when analysed together with morphological and behavioural traits, or with environmental variables, may shed light on the forces underlying species isolation and differentiation (Lexer, Buerkle, Joseph, Heinze, & Fay, 2007; Lindtke et al., 2012). Nonetheless, untangling the drivers of species differentiation and their genomic consequences can be difficult. The main reason is that hybrid zones tend to be positioned in areas of low population density and/or environmental ecotones (Endler, 1977; Leaché, Grummer, Harris, & Breckheimer, 2017; Taylor et al., 2014; Wielstra et al., 2017), even when environmental factors are not the main drivers of species isolation (Bierne, Welch, Loire, Bonhomme, & David, 2011). Besides, a hybrid zone associated with intrinsic reproductive isolation factors (e.g. genetic incompatibilities, assortative mating) may move and stabilize around an extrinsic natural barrier or ecotone, which may erroneously suggest a role for environmentally driven selection in the formation of the hybrid zone (Barton & Hewitt, 1985; Hewitt, 1988). In fact, the coupling of intrinsic and extrinsic isolating factors interacting along hybrid zones may be relatively common in the wild (Abbott et al., 2013; Bierne et al., 2011). Coincidence of barrier effects can occur either via selection on individual barrier effects or as a by-product of demographic processes. It can further evolve as an adaptive response to indirect selection, through the evolution of prezygotic barriers to gene flow in response to costs of hybridization (reinforcement) (Butlin & Smadja, 2017).

Here, we investigate the strength and dynamics of isolation in a contact zone between two alpine butterfly species in the genus *Coenonympha* (Nymphalidae, Satyrinae): the Alpine Heath *C.*

gardetta (Prunner 1798) and the Darwin Heath *C. macromma* (Turati & Verity 1910). *Coenonympha gardetta* is a typical alpine species, largely distributed at elevations above 1,500 m along the Alps, the French Massif Central and the Balkans, whereas *C. macromma* is restricted to the south part of the French Alps, where it replaces *C. gardetta*. The two species are found flying together in narrow zones where their distribution ranges abut. Hybrids are relatively common in these contact zones, which highlights incomplete reproductive isolation between species, even if they remain highly differentiated genetically at a large geographic scale (Capblancq, Mavárez, Rioux, & Després, 2019). Identifying the mechanisms involved in reproductive isolation between these species would provide a better understanding of the speciation process for *C. macromma*, which has been suggested to be the product of homoploid hybridization between *C. gardetta* and a third species, the Pearly Heath *C. arcania* (Linné 1761). Indeed, in a previous study (Capblancq, Després, Rioux, & Mavárez, 2015), we showed that *C. macromma* is composed of a genetic mixture between *C. gardetta* and *C. arcania* (30% and 70%, respectively) and that it exhibits intermediate wing morphology and ecological preferences (climatic niche). Statistical modelling of different past evolutionary scenarios allowed us to confirm the hybrid origin of *C. macromma* and estimate its emergence at around 12,000–20,000 years ago. Investigating the patterns of species divergence and its drivers in this clade would further provide more general insights into the process of homoploid hybrid speciation (Mavárez & Linares, 2008; Schumer, Rosenthal, & Andolfatto, 2014).

We used genomic, morphological and environmental data to investigate the fine-scale divergence between *C. gardetta* and *C. macromma*. After identifying an active and narrow hybrid zone between these two species, we performed a geographic cline analysis to characterize the genetic architecture of the hybrid zone and to investigate the genomic heterogeneity of introgression between the two species. We also looked for patterns of morphological and ecological divergence between species. Our goal was to evaluate the degree of concordance between ecological, morphological and genetic clines in the hybrid zone and to identify loci and traits potentially involved in species divergence and isolation.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We sampled 142 butterflies from 20 sites in the southern French Alps: 11 sites from a *C. gardetta/C. macromma* contact zone in which interspecific hybrids seem to be common and 9 sites located outside this area within the same region (Figure 1 and Table S1). The contact zone sites (P1 to P11; Figure 1) are distributed along an 8-km transect (P1 at position 0 m, P11 at position 8,100 m) located near the Col de Vars, a high-elevation alpine pass (2,100 m) situated where the ranges of *C. gardetta* (north) and *C. macromma* (south) overlap. Elevation and habitat change markedly along this transect, from mid-elevation semi-closed vegetation habitats (i.e. a high coverage

by shrubs and trees) at P1, to high-elevation open vegetation (grasslands) at P11. The 9 sites sampled outside the contact zone have individuals with wing traits characteristic of only one of either parental species (*C. gardetta*: LAU, SES, AIL, VJ; *C. macromma*: SEY, LAR, LOM, FOA, BOR; Figure 1 and Table S1). Sampled individuals were kept in glassine envelopes in the field and in ethanol 96% at -20°C in the laboratory until DNA extraction.

2.2 | Genetic data acquisition

DNA was extracted from the complete thorax of each individual using the DNeasy Blood and Tissue Kit (Qiagen, Germany). A data set of single nucleotide polymorphisms (SNPs) was produced using a double-digested restriction site-associated DNA (ddRAD) sequencing procedure using a modified version of the protocol of Peterson, Weber, Kay, Fisher, and Hoekstra (2012) as described in Capblancq et al. (2015). Five different *SbfI/MspI* ddRAD libraries were sequenced, each one in 1/10 of lane of a HiSeq 2,500 Illumina sequencer (Fasteris S.A., Switzerland). The obtained DNA reads (~100 million of 2×125 paired-end reads) were used to call SNP genotypes with the STACKS pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) using a Phred score of 10 for read filtering (*process_radtags* function), a minimum coverage of 4 to create a stack (*-m 4* in *ustacks* function) and a maximum of 6 different nucleotides to merge two different stacks (*-M 6*). Highly repetitive stacks and overmerged tags were dropped using both the "Removal algorithm" and the "Deleveraging algorithm." Furthermore, a maximum of 8 mismatches was allowed for considering two individual tags as the same locus and to merge them in the catalogue (*-n 8* in the *cstacks* function). In order to evaluate the experimental success for each library and each individual, we used *ProcessMyRAD* scripts (<https://github.com/cumtr/PmR>, Cumer et al., 2018) to produce graphical outputs at different steps of the treatment. This pipeline provides the proportion and number of reads retained after quality filtering, sequencing nucleotide quality along the reads, and the proportion of the different nucleotides. It also allows repeating the critical step of de novo individual locus assembly (*ustacks*) using a range of *M* values. By looking at the number of assembled RAD tags across *M* values, it provides a way to find a value allowing assembling a maximum of loci without stacking too many paralogs. *M = 6* was chosen because values above this threshold did not substantially increase the number of assembled loci within individuals (Figure S1).

To avoid an excess of missing data, all further analyses were performed using only RAD fragments genotyped for more than 40% of the samples of each of the three types of population (pure *C. gardetta*, pure *C. macromma* and contact zone) (parameters *-r 0.4* and *-p 3* in the *populations* function of STACKS). Only one polymorphic site was randomly kept for each polymorphic RAD fragment to avoid inclusion of physically linked SNPs. Finally, SNPs with a minor allele frequency lower than 1% (<2 individuals) were removed from the data set.

2.3 | Genetic diversity and structure across the populations

Genetic diversity statistics, that is observed heterozygosity (H_{obs}), gene diversity (H_{exp}) and *Fis*, were calculated for populations with at least 5 samples, using the function "basic.stats" of the *hierfstat* R package (Goudet & Jombart,). Population structure across the whole study area was assessed using sNMF, a constrained method of genetic clustering (see Frichot, Mathieu, Trouillon, Bouchard, & François, 2014). This analysis was performed using the R package *LEA* (Frichot & François, 2015), imposing two genetic clusters ($K = 2$) in order to estimate the probability of assignment of each individual to the two parental species genetic groups (*C. macromma* and *C. gardetta*). We assumed that the pure parental genetic makeup was correctly inferred in the analysis by using numerous (20 *C. macromma* and 39 *C. gardetta*) individuals from species-specific allopatric populations (Figure 1).

2.4 | Individual index of hybridization and interspecific heterozygosity estimation across the contact zone

Genetic admixture was estimated by maximum likelihood for each individual within the contact zone using the procedure implemented in the R package *introgress* (Gompert & Alex Buerkle, 2010). This procedure provides a genetic hybrid index (HINDEX) representing an estimate of the proportion of alleles inherited from one of the two parental species (Buerkle, 2005). The HINDEX value thus estimates a global index of admixture taking all the loci into account. Reference values for allele frequencies were established for each parental species from pure populations of *C. macromma* or *C. gardetta* (Figure 1 and Table S1). The contact zone was analysed considering all individuals as potential hybrids between the two parental allele frequency references. The HINDEX ranges from 0 to 1, with extreme values corresponding to pure individuals of each reference parental species. This method allows for the use of codominant markers and, quite appropriately for closely related species, for markers that are not necessarily fixed between taxa (Buerkle, 2005).

Interspecific heterozygosity was also estimated for each individual within the contact zone by using the function *calc.intersp.het* of the *introgress* R package, which calculates the proportion of an individual's genome with alleles inherited from each parental species (Buerkle, 2005). It gives an estimation of the direct bi-ancestrality of each sampled genotype. This method specifically estimates the heterozygosity due to admixture between the two defined parental groups, meaning that an F1 individual would return a value close to 1 (at least > 0.85 according to Milne & Abbott, 2008). Interspecific heterozygosity is thus designed to detect recent hybrids (e.g. F1, F2, backcrosses), as they are expected to show index values higher than parental individuals and ancient hybrids. Indeed, in populations where admixture took place a long

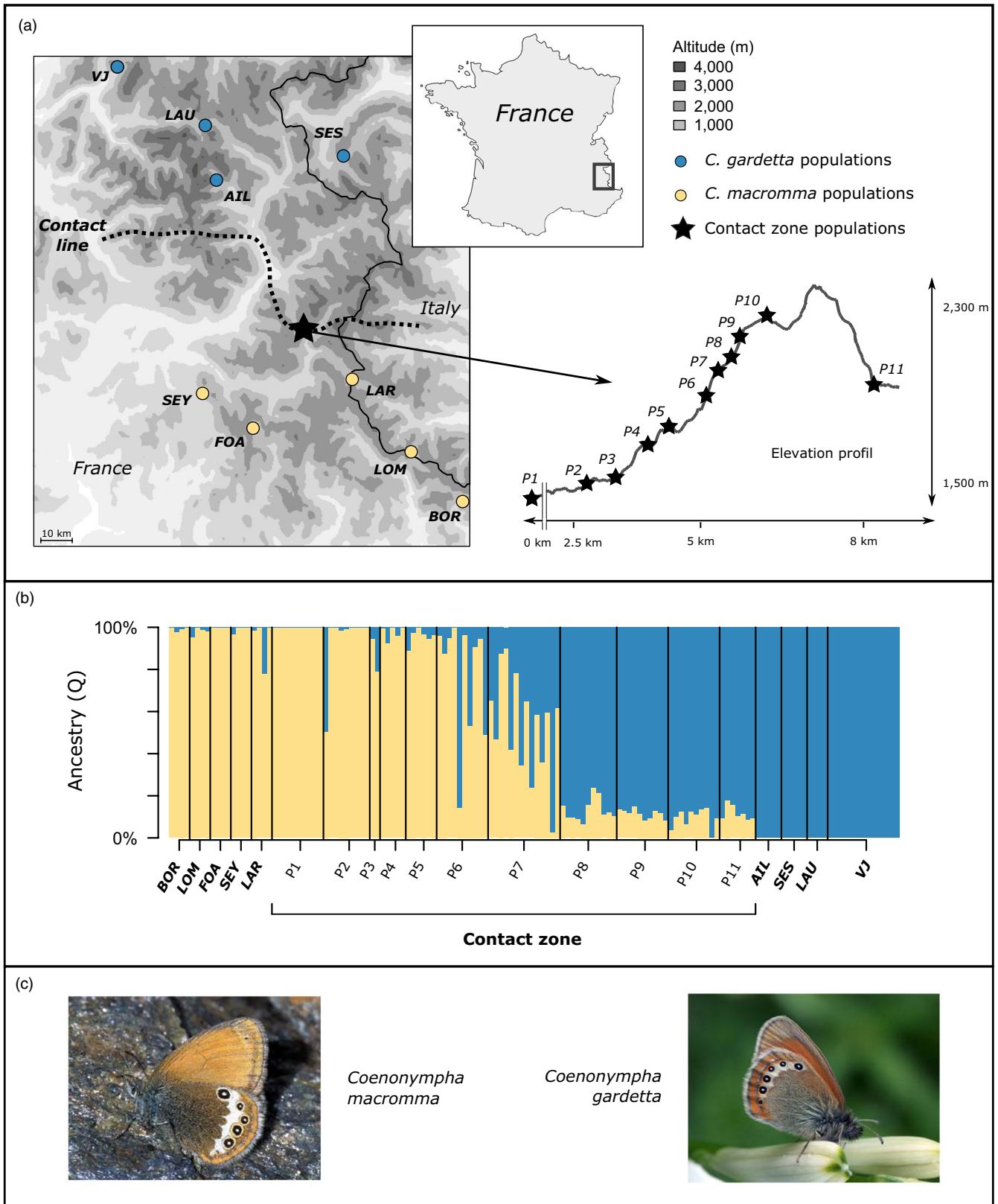


FIGURE 1 (a) Map showing the contact line between *Coenonympha macromma* and *C. gardetta* ranges in the French Alps, the location of sampled populations and the profile of the contact zone. (b) sNMF plot showing ancestry scores for all the individuals at $K = 2$, indicating genetic contributions from *C. macromma* or *C. gardetta* genetic clusters across the sampling zone. (c) Pictures of the two parental species, *C. gardetta* (Photo by Claire Hoddé, <https://www.lepinet.fr>) and *C. macromma* (Photo by Tristan Lafranchis, <https://www.lepinet.fr>)

time ago and with no contemporary interspecific gene flow, individuals are expected to show levels of interspecific heterozygosity similar to parental populations, but with a mosaic of alleles from either reference population (De La Torre et al., 2015; Fitzpatrick, 2012; Lindtke et al., 2012).

2.5 | Genetic clines along the contact zone

To further investigate the genetic transition between *C. macromma* and *C. gardetta* along the contact zone, we fitted a genetic cline model to the loci with differences in allele frequencies between parental populations >0.8 , using the *hzar* R package (Derryberry et al., 2014). We only analysed those highly divergent loci because cline analysis aims at looking at the change along the contact zone of features (genes or traits) characteristic of each parental population. We found that 122 SNP (hereafter “diagnostic loci”) fulfilled the minimum difference condition (see Results and Figure S2). The clines for the diagnostic loci were fitted with the 15 models designed by Derryberry et al. (2014). These 15 models are combinations of three possible allele frequency intervals ($p_{\min} - p_{\max}$ set to the observed values, fixed to 0 and 1, or estimated during the procedure) and five possible ways to fit tails (none fitted, left only, right only, mirror tails and both tails estimated separately). To ensure that the likelihood surface was correctly explored, we conducted three independent runs for each model, each run consisting of 20,000 iterations following an initial burn-in phase of 500 iterations. For each locus, we retained the model with the lowest Akaike information criterion (AIC) and extracted from this model the parameters for cline width (w) and centre (c) estimates. We also used the difference between the observed minimum and maximum allelic frequencies (ΔP) to estimate the slope (m) of each cline using the following equation: $m = (\Delta P/w)$ (Stankowski, Sobel, & Streisfeld, 2017).

2.6 | Morphological variations between parental species and along the contact zone

To identify whether the genetic transition between species, especially across the contact zone, was correlated with morphological variation, we performed geometric morphometric analyses of wing shape, patterns and venation. We focused on traits that are already known to discriminate between *C. macromma* and *C. gardetta*, such as wing size, the position of eyespots and the shape of the hindwing white band (see Capblancq et al., 2015). We used 22 and 18 homologous landmarks in the fore- and hind-wings, respectively, corresponding to vein intersections and termini, eyespot centres and the outline of the hindwing white band (Figure S3). The 80 coordinates resulting from these 40 landmarks were used to perform Procrustes superimposition in order to remove non-shape variation in location, scale and orientation. Our goal here was to identify different morphological traits that discriminated the two parental populations of *C. macromma* and *C. gardetta* and

then analyse the variation of these traits across the contact zone. We thus conducted a series of Procrustes superimposition on different sets of landmarks corresponding to four different traits: the vein intersections and termini of the forewing as a proxy of forewing global shape, the vein intersections and termini of the hindwing as a proxy of hindwing global shape, the eyespot landmarks as a proxy for eyespot positions, and the white band landmarks as a proxy for white band shape. We then performed for each trait a principal components analysis (PCA) on the superimposed coordinates and used the two first PCs to visualize and test for the potential divergence between parental populations of *C. macromma* and *C. gardetta*. We kept only the two first PCs for each trait because they gathered in each case a high proportion of the morphological variation (see Results), and we wanted to stay focused on the strongest interspecific morphological divergences between parental species. The trait differences between parental populations were tested using a one-way MANOVA on the first two PCs, with shape as the dependent variable and species as a factor. A wing size index was also obtained using the sum of the log-transformed centroid sizes estimated independently for the fore- and hindwing landmarks. Difference in size among species was tested with Student's *t* test with size as dependent variable and species as factor. The generalized Procrustes analysis and the different statistics were performed in R with the packages *Rmorph* (Baylac, 2012), *ade4* (Chessel, Dufour, & Thioulouse, 2004) and *APE* (Paradis, Claude, & Strimmer, 2014).

We then fitted seven cline models to the three morphological traits distinguishing parental species in the geometric morphometric analysis: wing size, the position of hindwing eyespots and the shape of the forewing white band (Figures S3 and S4). We fitted a cline model to these three traits with the *hzar* package and using the sum of the log-transformed centroid sizes for the fore- and hindwing as a proxy for the global wing size, the loadings of each individual along the first principal component (PC1) remaining from the Procrustes superimposition of the eyespot landmarks as a proxy for eyespot positions, and the loadings of each individual along the first principal component (PC1) remaining from the Procrustes superimposition of the white band landmarks as a proxy for white band shape. PC1 represented most of the variation for the two last traits (66 and 47%, respectively) and discriminated the parental populations (see Results). The seven models used were derived from the ones described for morphological traits in Derryberry et al. (2014) and are characterized by: a trait interval fixed to the observed values with five possible combinations of fitting tail (none fitted, left only, right only, mirror tails and both tails estimated separately) or a trait interval estimated during the procedure with two possible combinations of fitting tail (none fitted or both tails estimated separately). We then selected the best-fit model based on AIC and recovered from it the centre, width and slope of the cline.

The association between morphological traits and the individual genetic composition was tested through linear regression with HINDEX as response variable and either wing size, PC1 scores of eyespot alignment or PC1 scores of white band shape as explanatory

variable. All regressions were performed in R. The strength of the association represented by the r-squared of the linear regression was compared among traits.

2.7 | Characterization of ecological clines

To identify whether the genetic transition between species, especially across the contact zone, was correlated with environmental variation, we analysed ecological differences among populations using four climatic variables and three habitat variables. The four climatic variables were recovered from the WorldClim database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005, <http://worldclim.org>): annual mean temperature, annual precipitation, precipitation seasonality and temperature seasonality. Two habitat descriptors were obtained from the analysis of temporal series of NDVI (Normalized Difference Vegetation Index) rasters from MODIS satellite imagery every 16 days from 2000 to 2015 (Choler, 2015): the start date of the growing season and the maximum NDVI of the year, which is a proxy for primary productivity (Choler, 2015). The proportion of high vegetation structures (trees and shrubs) was extracted, for parental populations, from Corine Land Cover rasters of habitats (<http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012>) in a pixel of 250 m × 250 m around each sampled population. For the eleven localities within the contact zone, we re-estimated the proportion of high vegetation structures by sampling in each site three plots of 10 m² where at least one *Coenonympha* individual was found. On these plots, we counted the number of times a tree or a shrub is present at each of the 60 points distributed each 30 cm along the two diagonals. Difference in environmental conditions between species was tested with Student's *t* test with each of the environmental variables as dependent variable and species as factor, using the environmental values of each pure species locations: 5 sites for *C. macromma* and 4 sites for *C. gardetta* (Figure 1).

We fitted the same seven cline models as for morphological traits to the environmental variables showing a significant divergence between pure species populations (see above). Again, we selected the best-fit model based on AIC and recovered from it the cline centre, width and slope. The association between environmental variation and genetic composition of individuals was tested through independent linear regression with HINDEX as response variable and each environmental variable as explanatory variable. All the regressions were performed in R, and we recovered the strength of the association represented by the r-squared of the linear correlation to compare them among variables.

3 | RESULTS

3.1 | Genetic data

The ddRAD libraries produced ~ 100 million reads, 80% of which were used for further analyses owing to their high read quality (>30)

and uniform distribution of nucleotides along the read (Figure S5). These data allowed the assembly of 3,200 RAD fragments per individual on average, with a mean coverage of 35 reads/fragment for the 142 individuals analysed. The influence of the clustering parameter *M* on fragment reconstruction was visually inspected and guided the choice of *M* = 6 (Figure S1). A total of 899 RAD fragments were polymorphic, representing a total of 5,123 SNPs. However, retention of polymorphic RAD fragments that were present in at least 40% of the individuals in each of the three populations, and random selection of 1 SNP per fragment (*maf* > 0.01), resulted in a final data set of 794 SNPs.

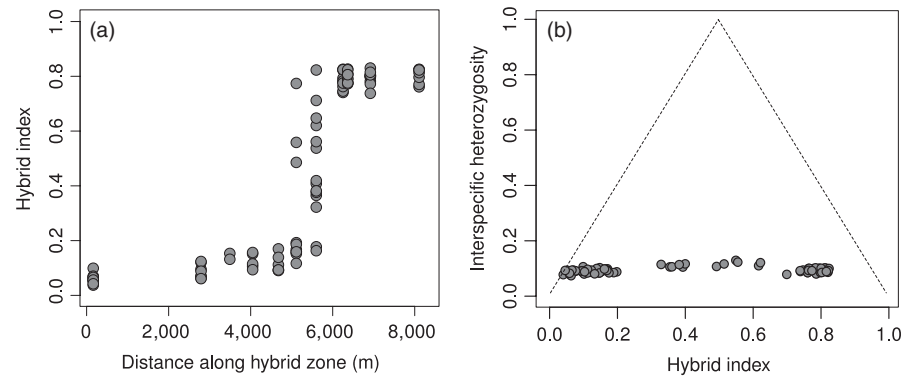
3.2 | Genetic characterization of the contact zone

The samples from the *C. macromma*/*C. gardetta* contact zone were highly admixed, whereas the allopatric populations of each species showed genetically pure individuals (Figure 1). The highest proportion of admixture was found in populations P6 and P7, in which almost only admixed individuals with variable ancestry proportions were collected. We observed a rapid decrease in the genetic contribution of *C. gardetta* from populations P1 to P5, with only genetically pure *C. macromma* individuals in P1 (Figure 1b). At the opposite, although the *C. macromma* genetic contribution strongly decreased from populations P8 to P11, a genetically pure *C. gardetta* was not observed, as individuals in P10 and P11 still retained a *C. macromma* genetic contribution of about 15%. The HINDEX analysis showed a similar pattern with estimates ranging from almost 0.06 in P1 to 0.84 in P11 and with intermediate values for P6 (0.3) and P7 (0.5) (Table 1). Furthermore, the variance of individual HINDEX estimates increased greatly within the two populations at the core of the hybrid zone (P6 and P7) in comparison with the other populations (Figure 2a).

TABLE 1 Table of sample size, genetic diversity estimates and mean HINDEX for the eleven populations sampled in the contact zone between *Coenonympha macromma* and *C. gardetta*

Population	<i>N</i>	<i>H</i> _{obs}	<i>H</i> _{exp}	<i>F</i> _{IS}	<i>H</i> _{index}
<i>C. macromma</i>	20	0.082	0.162	0.422	0
P1	10	0.082	0.153	0.352	0.06
P2	9	0.098	0.173	0.329	0.08
P3	2	–	–	–	0.14
P4	5	0.092	0.164	0.306	0.13
P5	6	0.102	0.171	0.306	0.12
P6	10	0.104	0.200	0.387	0.30
P7	14	0.118	0.197	0.340	0.50
P8	11	0.101	0.159	0.290	0.83
P9	10	0.102	0.160	0.294	0.81
P10	10	0.107	0.171	0.301	0.82
P11	7	0.098	0.161	0.298	0.84
<i>C. gardetta</i>	39	0.086	0.133	0.344	1

FIGURE 2 (a) Individual HINDEX estimates across the eleven populations constituting the contact zone between *Coenonympha macromma* and *C. gardetta* in the French Alps. (b) Interspecific heterozygosity (the fraction of markers heterozygous for alleles from both species) versus HINDEX



In parallel, we observed only a slight increase in genetic diversity (H_{exp}) and heterozygosity (H_{obs}) within the hybrid zone compared to pure populations of parental species (Table 1). The unexpected small increase in heterozygosity, considering the mixing of two divergent genetic backgrounds in this area, was confirmed by the *introgress* analysis, which returned a very low interspecific heterozygosity for the individuals sampled in the contact zone, even for the ones exhibiting the highest values of admixture (Figure 2b). The low levels of interspecific heterozygosity indicate that only a low proportion of loci per individual were inherited from the two parental populations. This means that the intermediate HINDEX values observed for many individuals in the contact zone were more likely due to a mosaic of parental contributions in the genome than to recent hybridization.

3.3 | Genetic cline analysis

The HINDEX cline, estimated from all 794 SNPs and representing the average genetic cline across the hybrid zone, exhibited a centre at 5,597 m, a width of 4,982 m, a slope value of 0.00015, near fixation on the left side and a value of 0.8 on the right (Figure 3 and Table S1). We observed a large variety of cline shapes for the 122 diagnostic loci (Figure 3). Most loci returned a relatively similar estimate for the centre of the cline, for example 60% ranging from 3,000 to 8,000 m. On the other hand, cline slope varied by nearly three orders of magnitude among loci, from almost 0 to 0.0027, with 66% of the loci exhibiting cline slopes higher than the HINDEX (Table S1). Cline width was strongly correlated with cline slope and varied from a few hundreds to several thousands of metres depending on the locus. Interestingly, we observed a strong relationship between centre and slope estimates, with the loci showing the sharpest clines being all centred at around 5,500–5,800 m (Figure 3).

The mean delta of cline tail values was relatively high ($\Delta P_{mean} = 0.88$), highlighting the fact that the 122 diagnostic loci, which are highly differentiated between the two species at large geographic scale, also show a strong turnover along the hybrid zone. A total of 56 of these loci even returned a cline with complete change in allelic frequency across the transect ($\Delta p = 1$), revealing that 46% of the species-specific alleles analysed here are

not exchanged between the two extremities of the hybrid zone. Furthermore, although 87% of the markers were fixed to 0 on the left side (*i.e.* “*C. macromma* side”) of the cline, only the fixed 56 reached 1 on the right tail (*i.e.* “*C. gardetta* side”). A subset of 23 loci exhibited a particularly sharp transition along the contact zone with an allelic frequency ranging from 0 to 1 in less than 3,000 m and a slope superior to 0.0005 (Figure 3). The consensus sequences of these loci were aligned against the ncbi nucleotide database using blastn, and ten of them matched Lepidoptera genes with high homology (>80%, E-value < 10^{-9} ; Table 2). Among these, we detected a putative olfactory receptor, and genes involved in cell regeneration processes, in oxidative stress and in eyespot formation.

3.4 | Variation in morphological traits

Out of the four morphological traits measured, three strongly differed between *C. macromma* and *C. gardetta*: the relative size of the wings (hind- and forewing centroid size combination, the *t* test returned a *p*-value < .001), the alignment of eyespots and the shape of the white band on the ventral side of the hindwing (MANOVA test returned a *p*-value < 2.2×10^{-16} for the two traits) (Figure S4). When analysing together the variation of the white band shape and the eyespot alignment in a PCA, we can observe a very strong differentiation between *C. macromma* and *C. gardetta* pure populations (Figure 4), with no overlap between individuals scores along PC1 (45% of variance explained). At the opposite, the populations from the contact zone exhibit a gradual range of scores along the same PC1, ranging from values close to pure *C. macromma* individuals for P1, P2 or P5, to values close to pure *C. gardetta* individuals for P8, P9, P10 and P11. Similar to the pattern observed for the genetic composition, populations P6 and P7 show very intermediate scores on the PCA performed with these two hindwing morphological traits. By contrast, the pattern of wing venation does not seem to differ between the two species, neither for the fore- nor for the hindwing (Figure S4).

Wing size returned a flat cline indicating random variations across the contact zone (Figure S6). In contrast, eyespot alignment, summarized by the individual scores along the first axis of a PCA performed on the eyespot centre landmarks (66% of variance

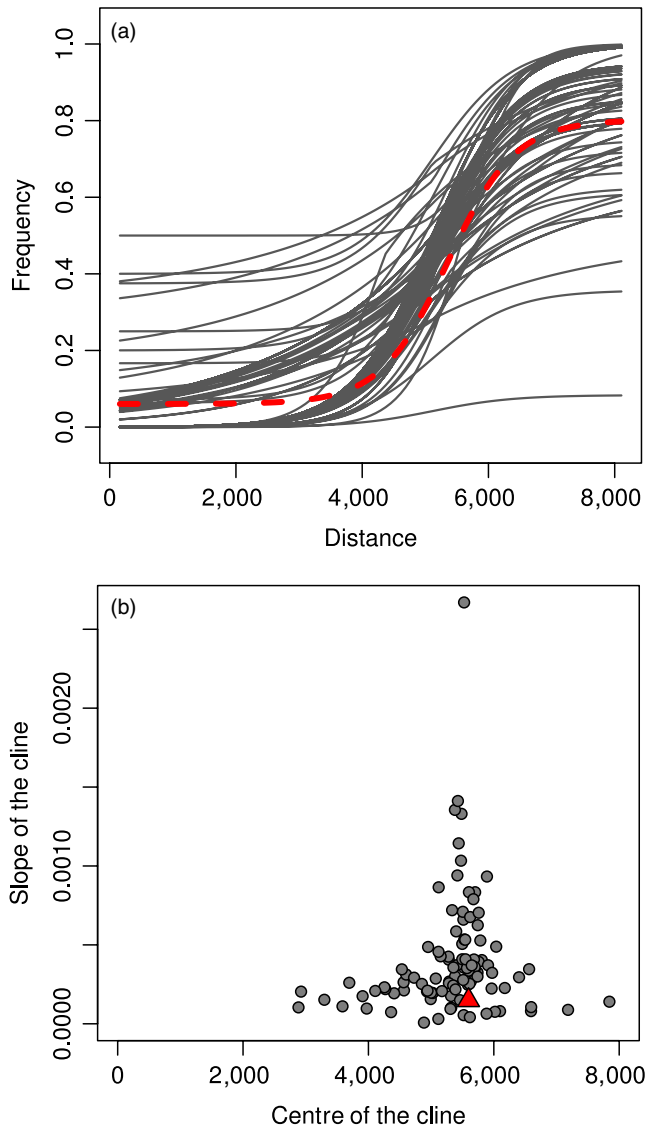


FIGURE 3 Plot of the maximum-likelihood cline for the best-fit model from the fifteen-model comparison for all the 122 species-specific loci (a). The red trait represents the HINDEX maximum-likelihood cline. (b) Relationship between cline centre and cline slope estimates from the ML cline of each locus. The red triangle represents the HINDEX centre and slope combination

explained), and white band shape, summarized by the individual scores along the first axis of a PCA performed on the white band landmarks (47% of variance explained), showed cline variations that coincide with the estimated centre of the genetic cline (Table 3 and Figure 5). Cline centre was located around 5,940 m for eyespots and around 5,260 m for the white band. The ML estimate for cline width was larger for eyespots (8 km) than for the white band (1.1 km). The linear regression between individual HINDEX estimates and morphological traits confirmed the association between genetic and morphological variation along the hybrid zone, pointing out to a stronger association between individual genetic composition and variation in the white band ($r^2 = .46$) than for eyespots ($r^2 = .25$).

3.5 | Pattern of ecological transition along the hybrid zone

Most of the environmental variables tested showed a significant difference between the locations of the pure populations of both *C. macromma* and *C. gardetta* (Figure S7). Only the maximum NDVI, a proxy for vegetation productivity, was not significantly different between species-specific locations, and elevation was only marginally significant (p -value .04). Furthermore, all the environmental variables that distinguish the two parental species showed a strong cline across the contact zone, with ML estimates of cline centre close to those obtained for genetic markers and morphological traits (4,300 m for annual mean temperature, 5,480 m for the proportion of trees and shrubs) (Figure 5 and Figure S8). The width of the clines varied from only 350 metres for the proportion of trees and shrubs to more than 2 km for the annual mean temperature (Table 3). The proportion of trees and shrubs, a variable characteristic of the habitat, showed the sharpest transition along the hybrid zone and returned a cline centre estimate (5,480 m) very close to the one obtained for the HINDEX (5,597 m). In the same way, the strongest statistical association was found between individual genetic composition (HINDEX) and the proportion of high vegetation structure along the hybrid zone ($r^2 = .9$; Table 3).

4 | DISCUSSION

4.1 | An active, narrow and long-lasting hybrid zone

Our study confirmed the presence of an active hybrid zone between *C. macromma* and *C. gardetta* in the south of the French Alps (near Col de Vars, Hautes-Alpes, France). We characterized in this area a set of 11 populations showing a gradient of admixture between pure *C. macromma* individuals and nearly pure *C. gardetta* individuals (Figures 1 and 2). Reproductive isolation is clearly not complete between these species, in agreement with a previous study, showing that the two taxa had undergone an uninterrupted history of contact and gene flow since their initial divergence 12,000–20,000 years ago (Capblancq et al., 2019). Yet, the two species keep their genetic integrities at large scale and the admixture found in this study remains restricted to the centre of a narrow hybrid zone (Figure 1; Capblancq et al., 2019).

The SNP cline analysis revealed a rapid genetic transition along a hybrid zone less than 9 km wide (Figure 3). Such a sharp genetic transition could reflect a recent secondary contact between the two species (i.e. ephemeral-zone hypothesis; Moore, 1977), but the low level of interspecific heterozygosity in hybrids does not support this hypothesis and rather suggests an advanced-generation hybrid zone (Figure 2). Simulation-based and empirical studies have shown that a hybrid zone with various degrees of individual admixture combined with low interspecific heterozygosity results from a long-lasting contact and widespread recombination between species (De La Torre et al., 2015; Fitzpatrick, 2012; Lindtke et al., 2012). Although RAD-like

TABLE 2 Characteristic of the genes strongly differentiated between *Coenonympha macromma* and *C. gardetta* that show the sharpest clines across the hybrid zone and that match other lepidopteran genes in ncbi database

Locus ID	Organism	Gene (predicted)	Putative function	Locus accession	% identity	E-value
CLocus_1159	<i>Papilio machaon</i>	G protein-coupled Mth 2-like	Olfactory receptor	XM_014503013.1	83.824	8.70E-10
CLocus_1160	<i>Galleria mellona</i>	Baculoviral IAP repeat-containing protein 6 like (BIRC6)	Immunity	XM_026902591.1	83.333	2.66E-16
CLocus_5199	<i>Bicyclus anynana</i>	Protein AF-10	Eyespot development	XM_024081740.1	84.783	6.26E-18
CLocus_4520	<i>Bicyclus anynana</i>	Secreted protein acidic rich in cysteine (SPARC)	Immunity	XM_024088674.1	94.118	3.69E-27
CLocus_3650	<i>Bicyclus anynana</i>	Rapamycin-insensitive companion of mTOR (RICTOR)	Cell growth regulation	XM_024085403.1	83.333	1.79E-18
CLocus_3620	<i>Bicyclus anynana</i>	Calcium-binding mitochondrial carrier protein SCaMC-2	Cell protection against oxidative stress	XM_013333945.1	85.714	2.66E-16
CLocus_1941	<i>Vanessa tameamea</i>	SET and MYND domain-containing protein 4-like	Cell growth regulation	XM_026644751.1	82.609	3.04E-09
CLocus_1630	<i>Papilio polytes</i>	Spatacsin	Larval colour pattern	XM_013290726.1	84.783	6.26E-18
CLocus_102796	<i>Bicyclus anynana</i>	Pre-mRNA cleavage complex 2 protein Pcf11-like	Transcription regulation	XM_024086011.1	97.297	1.57E-25
CLocus_112938	<i>Bicyclus anynana</i>	exocyst complex component 4-like	Exocytosis	XM_024082099.1	89.474	7.14E-11

Note: The percentage of homology, the putative function and the species for which these genes have been previously identified are shown.

experiments can underestimate heterozygosity (Cariou, Duret, & Charlat, 2016) and could have biased our results to a certain extent, such experimental bias cannot explain the very low level of interspecific heterozygosity found in the contact zone (Figure 2) or the small difference between parental and contact zone populations' heterozygosity values (Table 1). In a previous study, we found high levels of interspecific heterozygosity for individuals of the same complex of species, genotyped with the same ddRADseq experiment and analysed the same way (Capblancq et al., 2019), showing that the low interspecific heterozygosity found in the hybrid zone cannot be only due to experimental and/or bioinformatics treatments.

4.2 | Genetic architecture of introgression between *C. macromma* and *C. gardetta*

Hybrid zones are expected to show variation in introgression rates along the genome, with reduced introgression for genomic regions involved in disruptive selection, whereas neutral regions can be exchanged freely between species (Barton & Hewitt, 1985; Nosil, Funk, & Ortiz-Barrientos, 2009; Rieseberg et al., 1999). The regions displaying restricted circulation across the hybrid zone are therefore expected to be under disruptive selection and likely correspond to genomic regions responsible for genetic incompatibilities, prezygotic isolation or involved in local adaptation to differences in the

environment (Gompert, Lucas, Fordyce, Forister, & Nice, 2010; Walsh, Gregory Shriver, Olsen, & Kovach, 2016). Thus, the investigation of the genetic architecture of hybrid zones provides a way to estimate the nature of the selective agents driving the divergence between species and shed light on their patterns and strengths along the genome and across the landscape (De La Torre et al., 2015; Ravinet et al., 2017; Stankowski et al., 2017).

In the hybrid zone studied here, patterns of variation in the estimates of cline centre and slope revealed some heterogeneity in cline shapes. Cline slope estimates are highly variable among the 122 diagnostic loci, although most loci show slope estimates below the global genetic cline value (HINDEX), whereas a few show extremely strong values (Figure 3b). In the same way, the centre of the cline estimates can vary greatly along the hybrid zone, but the steepest clines are all centred around 5,600 m (Table S2). This kind of heterogeneous genomic divergence has already been described for many other species pairs (Fitzpatrick et al., 2009; Payseur, 2010; Payseur, Krenz, & Nachman, 2004; Rieseberg et al., 1999), especially when the process of speciation began recently (Nosil et al., 2009). Given that *C. macromma* and *C. gardetta* have likely been diverging for only a few thousands of years (12,000–20,000 years; Capblancq et al., 2015), it is not surprising to find such pattern in this hybrid zone, even if a quantitative association between divergence time and strength of isolation is not trivial when we deal with hybrid species (Mallet, 2007). Indeed, due to the reshuffling of parental

traits and genomes, even recent hybrid species can be strongly isolated from parents (Lukhtanov, Shapoval, Anokhin, Saifitdinova, & Kuznetsova, 2015; Nice et al., 2013; Schwander, Suni, Cahan, & Keller, 2008; Schwarz, Shoemaker, Botteri, & McPherson, 2007). In agreement with this assertion, there is probably a nearly complete reproductive isolation between *C. macromma* and its other parental species, *C. arcania*, since only one natural hybrid has been found

in a previously studied contact zone between these two species (Capblancq et al., 2019), although, according to the hybrid speciation scenario, they would have diverged for the same number of generations (12,000–20,000 years/generations).

Along the hybrid zone, some markers do not reach a $P_{\max} = 1$ in the right tail of their clines (i.e. "*C. gardetta* side"), even when scored several kilometres away from the centre (P11 in Figure 3a). At the opposite, almost all the markers return a P_{\min} of 0 in the left side of the cline (i.e. "*C. macromma* side"). It suggests that a part of *C. macromma* genetic material can introgress more easily into the *C. gardetta* genetic background than the reverse. Such asymmetry in the introgression pattern has already been associated with differences in population abundance (Lepais et al., 2009), with hybrid zone movement due to climate change (Taylor et al., 2014), or with scenarios in which the introgressive population is expanding while the introgressed population is declining (Johnson, White, Phillips, & Zamudio, 2015). A deeper investigation would be necessary to understand the causes of this pattern in the studied hybrid zone, although abundance does not seem to strongly differ between species in the sampled localities (field observations).

Almost half of the 122 diagnostic loci between *C. macromma* and *C. gardetta* seem to experience a restricted allelic transfer across the hybrid zone: the mean allelic frequency change along the hybrid zone is 0.88, with 46% of the loci even showing a complete frequency change (Figure 3 and Table S1). These 122 diagnostic loci also show cline width estimates not exceeding the spatial extent of the hybrid zone. Altogether, these results suggest that a portion of the genome is nearly impermeable to gene flow between the two species and that recombination of these genomic regions is restricted to the core of the hybrid zone. The degree of reproductive isolation between the species, if not complete, would thus be at least relatively strong,

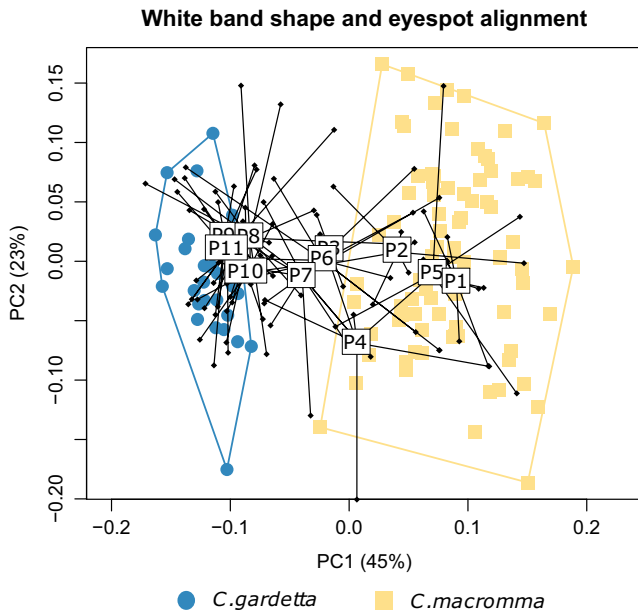


FIGURE 4 Morphological differentiation based on geometric morphometric analysis of two hindwing traits: the white band shape and the eyespot alignment. The individuals from the contact zone are shown by the black dots, and a box with the name of the population shows the barycentre of each of the 11 populations

TABLE 3 Names and maximum-likelihood parameter estimates of the centre, width and slope of the best-fit cline model for HINDEX, morphological traits and environmental variables

	Cline model	Centre	Width	Slope	<i>p</i> -value	<i>R</i> ²
Genetic						
Hindex	Free bounds and none fitted tails	5,597	4,982	0.00015	–	–
Morphology						
Eyespots	Free bounds and none fitted tails	5,938	8,465	0.00003	3.6e–05	0.25
White band	Free bounds and none fitted tails	5,262	1,174	0.00033	1.3e–06	0.46
Environment						
Altitude	Free bounds and none fitted tails	4,566	690	0.00099	<2e–16	0.71
Annual mean T _j C	Free bounds and both tails estimated separately	4,520	2,357	0.00036	<2e–16	0.69
Start of the growing season	Free bounds and none fitted tails	4,751	1,604	0.00051	<2e–16	0.74
Precipitation seasonality	Free bounds and both tails estimated separately	4,591	1,612	0.00125	9.8e–16	0.52
T _j C seasonality	Free bounds and both tails estimated separately	4,366	1,918	0.00043	<2e–16	0.67
Nb of Tree & Shrubs	Free bounds and none fitted tails	5,482	358	0.00222	<2e–16	0.9

Note: The last two columns give the *p*-value and *r*-squared (*R*²) of a linear regression between the HINDEX and the morphological traits (sample size: *n* = 81 individuals), and between the HINDEX and the environmental variables (*n* = 93 individuals) across the hybrid zone.

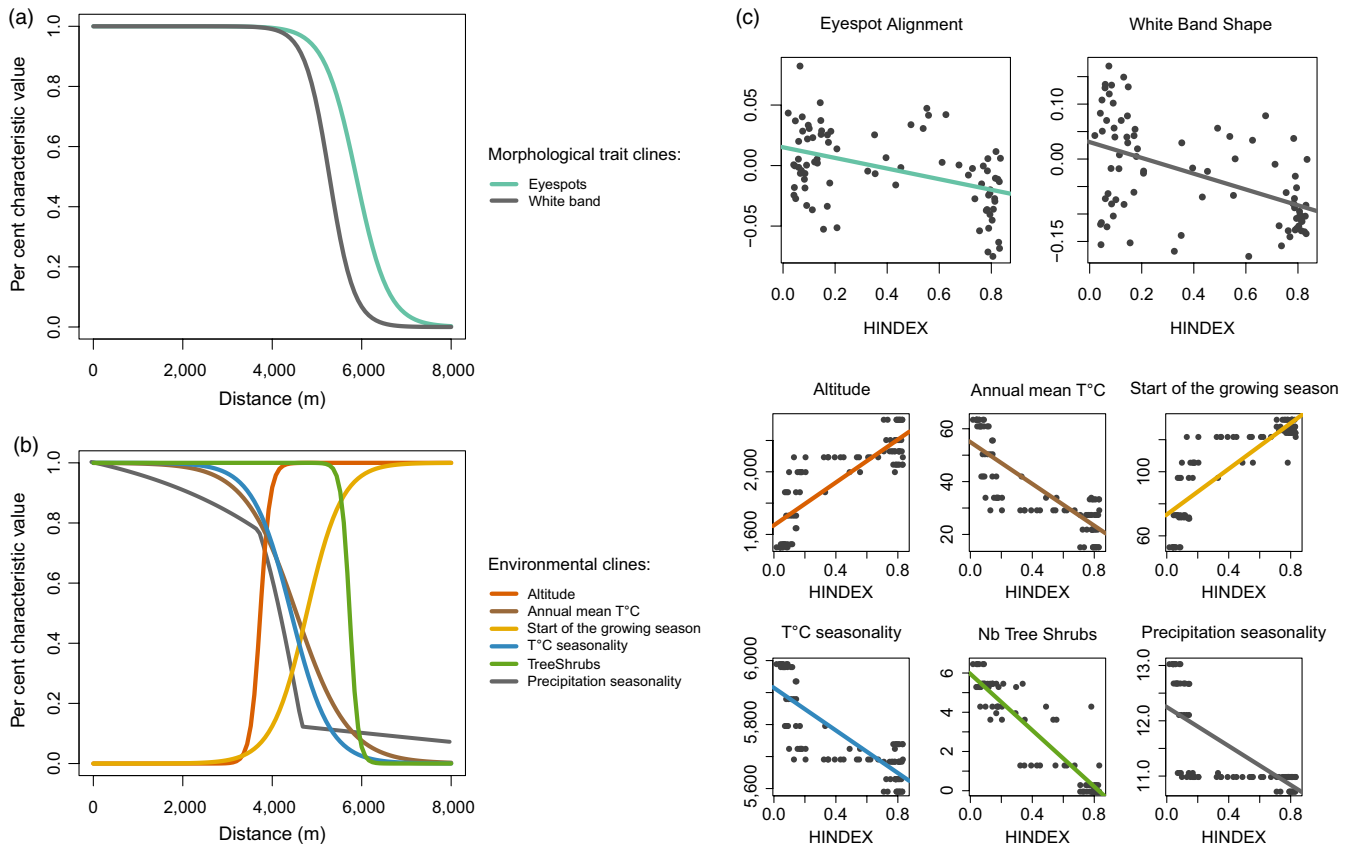


FIGURE 5 Plot of the per cent characteristic value versus distance for two morphological traits (a) and 5 environmental variables (b) describing the sampled hybrid zone. Each line is the maximum-likelihood fundamental cline for the best-fit model from the seven-model comparison. Panel (c) shows the variation of the different morphological and environmental traits along the HINDEX values with the lines showing the trends predicted by the associated linear regressions (p -value and R^2 in Table 3)

confirming the genetic integrity of species at a large geographic scale (Capblancq et al., 2015, 2019).

4.3 | Drivers of species isolation

The abrupt geographic transition observed between the genetic makeup of the two species and the unimodal distribution of admixed genotypes in the core of the hybrid zone could be consistent with two types of described hybrid zone models: a tension zone model (Buggs, 2007; Jiggins & Mallet, 2000; Wielstra et al., 2017) or a bounded hybrid superiority model (Abbott, 2017; Moore, 1977). In the tension zone theory, hybrid zones are maintained by a balance between species dispersal and selection against hybrids due to genetic incompatibilities between parental genomes (Barton, 2001; Barton & Hewitt, 1985). The sharpness of the genetic clines seen here would thus reflect strong selective constraints and/or low dispersal capacity of the species (Arias et al., 2012). Again, the low interspecific heterozygosity observed in this hybrid zone suggests a long-term and “stabilized” hybrid zone and, together with its narrow width, makes it unlikely that a poor dispersal capacity alone could prevent the complete mixing of the genomes in the area (Barton & Hewitt, 1989). Besides, it is unlikely that hybrids show a strongly

decreased fitness given their abundance in the core of the hybrid zone (field observation). On the other hand, the bounded hybrid superiority model states that strong disruptive selection mediated by behavioural or environmental factors could participate to shape hybrid zone dynamics (Abbott, 2017; Moore, 1977). In this case, hybrids would be more fit than parents in the core of the hybrid zone but restricted to this particular ecotone (Grant, Grant, Keller, Markert, & Petren, 2003; Mallet, 2005). Considering the number of diagnostic loci showing a very steep cline along the hybrid zone (56/122), it would suggest that various regions of the genome are involved in environmental disruptive selection. However, genomic regions involved in ecological adaptation are usually restrained to few parts of the genome (Nosil et al., 2009), although polygeny and epistasis can also be involved in ecological divergence (Albert et al., 2008; Arnegard et al., 2014; Des Roches, Sollmann, Calhoun, Rothstein, & Rosenblum, 2017).

Here, our results suggest a more complex interplay of different types of barriers to reproduction, partially associated with the particularity of the hybrid origin of *C. macromma*. We found a great concordance between genetic, morphological and environmental variations along the hybrid zone (Figure 3 and Figure 5), as often observed in natural hybrid zones involving recently diverged animal species (Arias et al., 2012; Rosser, Dasmahapatra, & Mallet, 2014;

Taylor et al., 2014). Both intrinsic barriers (e.g. assortative mating, habitat choice) and extrinsic barriers (e.g. local adaptation) could be involved in the reproductive isolation between *C. macromma* and *C. gardetta*. Among the loci least permeable to introgression in the hybrid zone, some are homologous to genes involved in olfactory receptors, eyespot development, immunity and cell protection against oxidative stress in other butterfly species. The more divergent genomic regions between *C. macromma* and *C. gardetta* may thus include genes putatively involved in assortative mating through wing pattern or pheromone recognition, or adaptations to stressful environmental conditions.

4.3.1 | Assortative mating

Highly visible wing patterns such as the white band and the eyespots of the ventral side of the hind-wings had steeper clines and were far more concordant with the genetic clines than wing size or venation patterns (Figure 5). This suggests the action of divergent selection on wing patterns that are visually different between the species, probably as a consequence of visual mate choice favouring assortative mating. In agreement with this hypothesis, one of the genomic regions showing the lowest introgression rate between the two species corresponded to a gene involved in eyespot development in *Bicyclus anynana* (Özsu, Chan, Chen, Gupta, & Monteiro, 2017). Further support for assortative mating as a source of isolation in this species pair comes from the observation that another genomic region with very low introgression rate corresponded to a putative olfactory receptor in *Papilio machaon*, suggesting that volatile organic compounds might be involved in chemical signalling and mate choice in *Coenonympha*. The involvement of both visual and olfactory cues in partner choice and species isolation has been demonstrated in many nymphalid butterflies (Costanzo & Monteiro, 2007; Mallet, Owen McMillan, & Jiggins, 2006; Mavárez et al., 2006; Mérot, Frérot, Leppik, & Joron, 2015; Pinzari et al., 2018). Although indirect, our results suggest that both visual and chemical signals might be involved in *Coenonympha* species reproductive isolation.

4.3.2 | Ecological divergence

Our results show that *C. gardetta* individuals are mostly found in open alpine grasslands, while *C. macromma* tend to be found in mixed habitats in which trees and large shrubs cover a significant fraction of the area. Indeed, the amount of high vegetation is the environmental factor most strongly associated with clinal genetic variation in this hybrid zone (Table 3). Alpine species have not only to handle with harsh climatic conditions including snow, low temperatures and storms, but they also have to face more wind, solar radiation, rapid desiccation and ample daily temperature variation in the absence of protective vegetation structures (Huey, 1991; Sinclair, Roger Worland, & Wharton, 2003; Turlure, Choutt, Baguette, Hans, & Dyck., 2010). This could explain some of the

ecological differences between *C. macromma* and *C. gardetta*, which are both adapted to high-elevation conditions but do not seem to use the same local habitats. Here again, several genetic markers found to be highly species-specific across the hybrid zone could be involved in this environmental-divergent adaptation, including a gene involved in cell oxidative stress, which is supposedly higher at full solar radiation than under vegetation cover (Meng, Zhang, Zhu, Wang, & Lei, 2009). Several genes involved in cell growth regulation/transcription might also be involved in differential physiological responses to stress-full conditions in open light (UV damages) as opposed to the possibility for *C. macromma* to behaviourally avoid these damages and overheating by hiding within high vegetation. In the same vein, it has been shown for another butterfly of the same genus, *C. pamphilus*, that vegetation structure, as a mating place, is key for reproductive success (Wickman, 1992). If a similar process drives habitat choice for *C. macromma*, the lack of high vegetation in high-elevation open meadows could be a strong barrier to hybridization with *C. gardetta*.

4.3.3 | Genetic incompatibilities

None of our results can confirm or reject the presence of genetic incompatibilities between *C. macromma* and *C. gardetta*. However, several genes showing low or no introgression between the two genomes involve basic general cellular processes such as cell growth or transcription regulation: these processes typically involve many interacting genes. Barrier effects in these genes can evolve through the spread of incompatible alleles via drift in the two diverging populations. Intrinsic incompatibilities could also evolve as a by-product of divergent ecological selection (Kulmuni & Westram, 2017). Indeed, genes can have multiple functions (pleiotropy) and/or function in complex networks, and their evolution is not independent of other genes: changes evolving in response to ecological selection can generate intrinsic barriers as a by-product.

Altogether, our results suggest that the abrupt geographic transition observed between the two genetic pools is shaped by the action of various ongoing selective pressure(s) that maintain the species integrity despite incomplete isolation and many generations of genetic exchange. We go a step further in dissecting the barrier effects, and found that both intrinsic (assortative mating through visual and olfactory cues) and extrinsic (vegetation structure) barriers are probably involved in reproductive isolation between *C. macromma* and *C. gardetta*, with various genomic regions involved, and several putative candidate genes are identified. The sharpness of the clines fitted for those genomic regions suggests ongoing selection. However, neutral markers can also exhibit very sharp clines under the effect of strong genetic drift increasing the probability of fixation of the major allele on both sides of the hybrid zone (Polechová & Barton, 2011). We cannot completely rule out this possibility here even if the strong concordance of the identified cline centres in the middle of the hybrid zone makes this pattern unlikely to be due to

drift alone (Figure 3). Further work would thus be needed to validate the involvement of visual and olfactory cues in mate choice in these species, and to evaluate the many possible impacts of habitat structure not only on mate choice, but also on the diversity of predators/pathogens and predator-escape strategies, and on the thermo-regulation strategies.

4.4 | What future for the two species in the Alps?

This study provides evidence for coincidence of morphological, ecological and genetic clines in a narrow hybrid zone where the overall barrier to gene flow is strong, but not complete. The incompleteness of the reproductive isolation observed in these two alpine taxa highlights the tenuous balance between dispersal and selection commonly observed in hybrid zones (Abbott, 2017; Barton & Hewitt, 1985, 1989; Buggs, 2007; Moore, 1977). The Alps are currently experiencing dramatic environmental changes, due to both the evolution of human activities and global warming (Lavergne, Mouquet, Thuiller, & Ronce, 2010). If, as we suggest here, the differentiation between *C. gardetta* and *C. macromma* is partially maintained by divergent environmental requirements, we might expect a disruption of the hybrid zone equilibrium in the next decades. The rise of temperature together with the increase in shrub/tree coverage at high elevations could favour *C. macromma* over its parental species *C. gardetta*, pushing the contact zone towards northernmost latitudes, as already reported for many other hybrid zones (Buggs, 2007; Ryan et al., 2018; Taylor et al., 2014; Thurman, Szejner-Sigal, & Owen McMillan, 2019; Wielstra et al., 2017). More dramatically, an unconstrained introgression of *C. macromma* genetic material into *C. gardetta* genetic pool would threaten the genetic integrity of the species in the future, at least locally (Hohenlohe, Amish, Catchen, Allendorf, & Luikart, 2011; Vonlanthen et al., 2012). Monitoring the potential change that will happen in this hybrid zone would thus be of critical importance to envision the future of this particular couple of species and could, more generally, act as a window on biodiversity threat in the Alps (Taylor, Larson, & Harrison, 2015; Wielstra, 2019). Our ability to predict such changes will be key to future conservation and management efforts.

ACKNOWLEDGEMENTS

We are very grateful to Émilie Crouzat, Aurélie Bonin and Iris Lochon for their help in field data collection and to Delphine Rioux for her help in the laboratory. We would also like to thank Marie-France Leccia and the Mercantour National Parc for helping us with sampling authorizations. We thank Nicolas Bierne and Marianne Elias for advice on the study and Mathieu Joron for helpful comments on the manuscript. This project was funded by the "Conseil Général de l'Isère" and a "CNRS INEE–APEGE" grant.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

J. Mavarez, L. Després and T. Capblancq designed the study and collected the samples. T. C. performed ddRAD experiments and bioinformatics treatment of the sequences, as well as the genetic, morphologic and environmental analyses. L.D. performed the ncbi blast analysis. T. C., J. M. and L. D. wrote the manuscript.

DATA AVAILABILITY STATEMENT

Genetic data in the form of a table of genotypes for each individual and each retained site are available at Dryad Digital Repository (<https://doi.org/10.5061/dryad.4mj8gg0>). Morphologic data in the form of the coordinates of the 40 morphometric landmarks for each individual are available at Dryad Digital Repository (<https://doi.org/10.5061/dryad.4mj8gg0>). Environmental data in the form of climatic rasters and a table of habitat variables are available at Dryad Digital Repository (<https://doi.org/10.5061/dryad.4mj8gg0>). The ProcessMyRAD pipeline is freely available on Github at <https://github.com/cumtr/PmR>. Scripts used to perform the different analyses are freely available on Github at <https://github.com/Capblancq/Speciation-Coenonympha-butterflies>. Digitalized pictures of the butterfly wings are freely available upon request from T. Capblancq.

ORCID

Thibaut Capblancq  <https://orcid.org/0000-0001-5024-1302>

Laurence Després  <https://orcid.org/0000-0002-0660-6260>

REFERENCES

- Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, 55(4), 238–258. <https://doi.org/10.1111/jse.12267>
- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and Speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Albert, A. Y. K., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., ... Schluter, D. (2008). The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution*, 62(1), 76–85.
- Arias, C. F., Rosales, C., Salazar, C., Castaño, J., Bermingham, E., Linares, M., & McMillan, W. O. (2012). Sharp genetic discontinuity across a unimodal heliconius hybrid zone. *Molecular Ecology*, 21(23), 5778–5794.
- Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., ... Schluter, D. (2014). Genetics of ecological divergence during speciation. *Nature*, 511(7509), 307–311. <https://doi.org/10.1038/nature13301>
- Barton, N. H. (1979). The dynamics of hybrid zones. *Heredity*, 43(3), 341–359. <https://doi.org/10.1038/hdy.1979.87>
- Barton, N. H. (2001). The Role of Hybridization in Evolution. *Molecular Ecology*, 10(3), 551–568. <https://doi.org/10.1046/j.1365-294x.2001.01216.x>
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, 16(1), 113–148. <https://doi.org/10.1146/annurev.es.16.110185.000553>
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, 341(6242), 497–503. <https://doi.org/10.1038/341497a0>
- 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. <https://doi.org/10.1111/j.1365-294X.2011.05080.x>
- Buerkle, C. A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, 5(3), 684–687. <https://doi.org/10.1111/j.1471-8286.2005.01011.x>
- Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, 99(3), 301–312. <https://doi.org/10.1038/sj.hdy.6800997>
- Butlin, R. K., & Smadja, C. M. (2017). Coupling, reinforcement, and speciation. *The American Naturalist*, 191(2), 155–172. <https://doi.org/10.1086/695136>
- Capblancq, T., Després, L., Rioux, D., & Mavárez, J. (2015). Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, 24(24), 6209–6222.
- Capblancq, T., Mavárez, J., Rioux, D., & Després, L. (2019). Speciation with Gene Flow: Evidence from a Complex of Alpine Butterflies (*Coenonympha*, Satyridae). *Ecology and Evolution*, 9, 6444–6457.
- Cariou, M., Duret, L., & Charlat, S. (2016). How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology*, 16(1), 1–8. <https://doi.org/10.1186/s12862-016-0791-0>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The Ade4 package—One-table methods. *R News*, 4(1), 5–10.
- Choler, P. (2015). Growth response of temperate mountain grasslands to inter-annual variations in snow cover duration. *Biogeosciences*, 12(12), 3885–3897. <https://doi.org/10.5194/bg-12-3885-2015>
- Costanzo, K., & Monteiro, A. (2007). The use of chemical and visual cues in female choice in the butterfly *Bicyclus Anynana*. *Proceedings of the Royal Society B: Biological Sciences*, 274(1611), 845–851.
- Cumer, T., Pouchon, C., Yannic, G., Rioux, D., Bonin, A., & Capblancq, T. (2018). Double-digest RAD-sequencing: Do wet and dry protocol parameters impact biological results? *BioRxiv*. <https://doi.org/10.1101/445122>
- De La Torre, A., Ingvarsson, P. K., & Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, 115(2), 153–164. <https://doi.org/10.1038/hdy.2015.19>
- Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). Hzar: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, 14(3), 652–663.
- Des Roches, S., Sollmann, R., Calhoun, K., Rothstein, A. P., & Rosenblum, E. B. (2017). Survival by genotype: Patterns at Mc1r are not black and white at the white sands ecotone. *Molecular Ecology*, 26(1), 320–329.
- Endler, J. A. (1977). *Geographic variation, speciation, and clines*. Princeton, NJ: Princeton University Press.
- Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, 12(1), 1–15. <https://doi.org/10.1186/1471-2148-12-131>
- Fitzpatrick, B. M., Johnson, J. R., Kevin Kump, D., Bradley Shaffer, H., Smith, J. J., & Randal Voss, S. (2009). Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. *BMC Evolutionary Biology*, 9(1), 176. <https://doi.org/10.1186/1471-2148-9-176>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983. <https://doi.org/10.1534/genetics.113.160572>
- Gay, L., Crochet, P.-A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, 62(11), 2789–2806. <https://doi.org/10.1111/j.1558-5646.2008.00491.x>
- Gompert, Z., & Alex Buerkle, C. (2010). Introgress: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, 10(2), 378–384. <https://doi.org/10.1111/j.1755-0998.2009.02733.x>
- Gompert, Z., Lucas, L. K., Fordyce, J. A., Forister, M. L., & Nice, C. C. (2010). Secondary contact between *Lycaeides idas* and *L. melissa* in the rocky mountains: Extensive admixture and a patchy hybrid zone. *Molecular Ecology*, 19(15), 3171–3192.
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186.
- Grant, P., Grant, R., Keller, L., Markert, J., & Petren, K. (2003). Inbreeding and interbreeding in Darwin's finches. *Evolution*, 57(12), 2911–2916. <https://doi.org/10.1111/j.0014-3820.2003.tb01532.x>
- Haldane, J. B. S. (1948). The theory of a cline. *Journal of Genetics*, 48, 277–284. <https://doi.org/10.1007/BF02986626>
- Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, 56(3), 337–349. <https://doi.org/10.1038/hdy.1986.55>
- Harrison, R. G. (1993). *Hybrid zones and the evolutionary process*. Oxford: Oxford University Press, 3–12.
- Hewitt, G. M. (1988). Hybrid zones—natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3(7), 158–167. [https://doi.org/10.1016/0169-5347\(88\)90033-X](https://doi.org/10.1016/0169-5347(88)90033-X)
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978. <https://doi.org/10.1002/joc.1276>
- Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridisation between rainbow and WESTSLOPE cutthroat trout. *Molecular Ecology Resources*, 11(March), 117–122.
- Huey, R. B. (1991). Physiological consequences of habitat selection. *The American Naturalist*, 137, 91–115. <https://doi.org/10.1086/285141>
- Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, 15(6), 250–255. [https://doi.org/10.1016/S0169-5347\(00\)01873-5](https://doi.org/10.1016/S0169-5347(00)01873-5)
- Johnson, B. B., White, T. A., Phillips, C. A., & Zamudio, K. R. (2015). Asymmetric introgression in a spotted salamander hybrid zone. *Journal of Heredity*, July, esv042.
- Kulmuni, J., & Westram, A. M. (2017). Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: considering them in empirical studies on divergence with gene flow. *Molecular Ecology*, 26(12), 3093–3103. <https://doi.org/10.1111/mec.14147>
- Lavergne, S., Mouquet, N., Thuiller, W., & Ronce, O. (2010). Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 321–350. <https://doi.org/10.1146/annurev-ecolsys-102209-144628>
- Leaché, A. D., Grummer, J. A., Harris, R. B., & Breckheimer, I. K. (2017). Evidence for concerted movement of nuclear and mitochondrial clines in a lizard hybrid zone. *Molecular Ecology*, 26(8), 2306–2316. <https://doi.org/10.1111/mec.14033>
- Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., & Gerber, S. (2009). Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, 18(10), 2228–2242. <https://doi.org/10.1111/j.1365-294X.2009.04137.x>
- Lexer, C., Buerkle, C. A., Joseph, J. A., Heinze, B., & Fay, M. F. (2007). Admixture in European populus hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, 98(2), 74–84. <https://doi.org/10.1038/sj.hdy.6800898>
- Lindtke, D., Buerkle, C. A., Barbará, T., Heinze, B., Castiglione, S., Bartha, D., & Lexer, C. (2012). Recombinant hybrids retain Heterozygosity at

- Many Loci: New insights into the genomics of reproductive isolation in *Populus*. *Molecular Ecology*, 21(20), 5042–5058.
- Lukhtanov, V. A., Shapoval, N. A., Anokhin, B. A., Saifitdinova, A. F., & Kuznetsova, V. G. (2015). Homoploid hybrid speciation and genome evolution via chromosome sorting. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1807), 20150157. <https://doi.org/10.1098/rspb.2015.0157>
- Mallet, J. (2005). Hybridisation as an Invasion of the Genome. *Trends in Ecology & Evolution*, 20(5), 229–237.
- Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279–283. <https://doi.org/10.1038/nature05706>
- Mallet, J., Owen McMillan, W., & Jiggins, C. D. (2006). Estimating the mating behavior of a pair of hybridizing *Heliconius* species in the wild. *Evolution*, 52(2), 503.
- Mavárez, J., & Linares, M. (2008). Homoploid hybrid speciation in animals. *Molecular Ecology*, 17(19), 4181–4185. <https://doi.org/10.1111/j.1365-294X.2008.03898.x>
- Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridisation in *Heliconius* butterflies. *Nature*, 441(7095), 868–871.
- Meng, J. Y., Zhang, C. Y., Zhu, F., Wang, X. P., & Lei, C. L. (2009). Ultraviolet light-induced oxidative stress: Effects on antioxidant response of *Helicoverpa armigera* adults. *Journal of Insect Physiology*, 55(6), 588–592. <https://doi.org/10.1016/j.jinsphys.2009.03.003>
- Mérot, C., Frérot, B., Leppik, E., & Joron, M. (2015). Beyond magic traits: Multimodal mating cues in *Heliconius* butterflies. *Evolution*, 69(11), 2891–2904.
- Milne, R. I., & Abbott, R. J. (2008). Reproductive isolation among two interfertile rhododendron species: low frequency of Post-F1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology*, 17(4), 1108–1121. <https://doi.org/10.1111/j.1365-294X.2007.03643.x>
- Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, 52(3), 263–277. <https://doi.org/10.1086/409995>
- Nice, C. C., Gompert, Z., Fordyce, J. A., Forister, M. L., Lucas, L. K., & Alex Buerkle, C. (2013). Hybrid speciation and independent evolution in lineages of alpine butterflies. *Evolution*, 67(4), 1055–1068. <https://doi.org/10.1111/evo.12019>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Özsu, N., Chan, Q. Y., Chen, B., Gupta, M. D., & Monteiro, A. (2017). Wingless is a positive regulator of eyespot color patterns in *Bicyclus anynana* butterflies. *Developmental Biology*, 429(1), 177–185. <https://doi.org/10.1016/j.ydbio.2017.06.030>
- Paradis, E., Claude, J., & Strimmer, K. (2014). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290. <http://cran.r-project.org/web/packages/ape/index.html>
- Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, 10(5), 806–820. <https://doi.org/10.1111/j.1755-0998.2010.02883.x>
- Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of introgression across the X Chromosome in a hybrid zone between two species of house mice. *Evolution*, 58(9), 2064–2078. <https://doi.org/10.1111/j.0014-3820.2004.tb00490.x>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Pinzari, M., Santonico, M., Pennazza, G., Martinelli, E., Capuano, R., Paolesse, R., ... Di Natale, C. (2018). Chemically mediated species recognition in two sympatric grayling butterflies: *Hipparchia Fagi* and *Hipparchia Hermione* (Lepidoptera: Nymphalidae, Satyrinae). *PLoS ONE*, 13(6), e0199997. <https://doi.org/10.1371/journal.pone.0199997>
- Polechová, J., & Barton, N. (2011). Genetic drift widens the expected cline but narrows the expected cline width. *Genetics*, 189(1), 227–235. <https://doi.org/10.1534/genetics.111.129817>
- 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. <https://doi.org/10.1111/jeb.13047>
- 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.
- Rosser, N., Dasmahapatra, K. K., & Mallet, J. (2014). Stable heliconius butterfly hybrid zones are correlated with a local rainfall peak at the edge of the amazon basin. *Evolution*, 68(12), 3470–3484.
- Ryan, S. F., Deines, J. M., Mark Scriber, J., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences*, 115(10), 201714950. <https://doi.org/10.1073/pnas.1714950115>
- Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How common is homoploid hybrid speciation? *Evolution*, 68(6), 1553–1560. <https://doi.org/10.1111/evo.12399>
- Schwander, T., Suni, S. S., Cahan, S. H., & Keller, L. (2008). Mechanisms of reproductive isolation between an ant species of hybrid origin and one of its parents. *Evolution*, 62(7), 1635–1643. <https://doi.org/10.1111/j.1558-5646.2008.00387.x>
- Schwarz, D., Shoemaker, K. D., Botteri, N. L., & McPheron, B. A. (2007). A novel preference for an invasive plant as a mechanism for animal hybrid speciation. *Evolution*, 61(2), 245–256. <https://doi.org/10.1111/j.1558-5646.2007.00027.x>
- Sinclair, B. J., Roger Worland, M., & Wharton, D. A. (2003). Ice nucleation and freezing tolerance in New Zealand alpine and lowland weta, *Hemideina* Spp. (Orthoptera; Stenopelmatidae). *Physiological Entomology*, 24(1), 56–63.
- Stankowski, S., Sobel, J. M., & Streisfeld, M. A. (2017). Geographic cline analysis as a tool for studying genome-wide variation: A case study of pollinator-mediated divergence in a monkeyflower. *Molecular Ecology*, 26(1), 107–122.
- Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: windows on climate change. *Trends in Ecology and Evolution*, 30(7), 398–406. <https://doi.org/10.1016/j.tree.2015.04.010>
- Taylor, S. A., White, T. A., Hochachka, W. M., Ferretti, V., Curry, R. L., & Lovette, I. (2014). Climate-mediated movement of an avian hybrid zone. *Current Biology*, 24(6), 671–676. <https://doi.org/10.1016/j.cub.2014.01.069>
- Thurman, T. J., Szejner-Sigal, A., & Owen McMillan, W. (2019). Movement of a heliconius hybrid zone over 30 years: A Bayesian approach. *Journal of Evolutionary Biology*, 32(9), 974–983.
- Turlure, C., Choutt, J., Baguette, M., Van Hans, A. N. S., & Dyck., (2010). Microclimatic buffering and resource-based habitat in a glacial relict butterfly: significance for conservation under climate change. *Global Change Biology*, 16(6), 1883–1893. <https://doi.org/10.1111/j.1365-2486.2009.02133.x>
- Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., ... Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482(7385), 357–362. <https://doi.org/10.1038/nature10824>
- Walsh, J., Gregory Shriver, W., Olsen, B. J., & Kovach, A. I. (2016). Differential introgression and the maintenance of species boundaries in an advanced generation avian hybrid zone. *BMC Evolutionary Biology*, 16, (1). <https://doi.org/10.1186/s12862-016-0635-y>
- Wickman, P. O. (1992). Mating systems of coenonympha butterflies in relation to longevity. *Animal Behaviour*, 44(1), 141–148. [https://doi.org/10.1016/S0003-3472\(05\)80763-8](https://doi.org/10.1016/S0003-3472(05)80763-8)

- Wielstra, B. (2019). Historical hybrid zone movement: more pervasive than appreciated. *Journal of Biogeography*, 46(7), 1300–1305. <https://doi.org/10.1111/jbi.13600>
- Wielstra, B., Burke, T., Butlin, R. K., Avcı, A., Üzümlü, N., Bozkurt, E., ... Arntzen, J. W. (2017). A genomic footprint of hybrid zone movement in crested newts. *Evolution Letters*, 1(2), 93–101. <https://doi.org/10.1002/evl3.9>

How to cite this article: Capblancq T, Després L, Mavárez J. Genetic, morphological and ecological variation across a sharp hybrid zone between two alpine butterfly species. *Evol Appl.* 2020;13:1435–1450. <https://doi.org/10.1111/eva.12925>

SUPPORTING INFORMATION

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