

# Evolutionary history and interspecific gene flow in six *Coenonympha* butterfly species in Europe

## Histoire évolutive et flux de gène entre six espèces de papillons du genre *Coenonympha* en Europe

Thibaud Camizuli<sup>1</sup>  | Matthew Greenwood<sup>1</sup>  | Delphine Rioux<sup>1</sup> |  
Elena Zakharova<sup>2</sup>  | Laurence Després<sup>1</sup> 

<sup>1</sup>Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, Université Savoie Mont-Blanc, CNRS, Grenoble, France

<sup>2</sup>Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Ekaterinburg, Russia

### Correspondence

Laurence Després, Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, 38058 Grenoble cedex 9, France.  
Email: [laurence.despres@univ-grenoble-alpes.fr](mailto:laurence.despres@univ-grenoble-alpes.fr)

### Funding information

Université Grenoble Alpes; Agence Nationale de la Recherche; French National Research Agency, Grant/Award Number: ANR-20-CE02-0017

Editor: Nusha Keyghobadi and Associate Editor: Marta Vila

### Abstract

1. Recently diverged species often exhibit incomplete reproductive barriers, leading to potential hybridisation where their ranges overlap. Distinguishing the effects of past hybridisation and incomplete lineage sorting from contemporary secondary gene flow in species evolution remains challenging, yet it is critical in conservation biology.
2. To investigate species boundaries and the diversification history of six closely related *Coenonympha* butterflies, we analysed their genomic diversity using ddRAD-seq data from 140 individuals sampled in two broad geographical regions (the Alps and the Southern Urals) where species distributions overlap. We tested for hybridisation amongst species and contrasted various divergence scenarios, incorporating gene flow and hybrid speciation.
3. In addition to the known gene flow history between the pearly heath *C. arcania* and the alpine heath *C. gardetta*, which led to the formation of two currently allopatric hybrid species (*C. darwiniana* and *C. cephalidarwinia*) in the Alps, we identified signatures of post-speciation gene flow between *C. arcania* and both the Russian heath *C. leander* and the scarce heath *C. hero* at a period of increased amplitude of the Pleistocene climatic oscillations, around 800,000 years ago. Amongst these species, the least genetically diverse is *C. hero*, which is the only species protected by a high conservation status.
4. Our findings reveal widespread secondary gene flow within this butterfly group, yet the six species remain well-differentiated genetically and morphologically, even in regions where their distributions overlap, suggesting ancient rather than current

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Insect Conservation and Diversity* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society.

gene flow. This highlights the need for broadscale studies to better understand the biogeographical and ecological context of species diversification.

#### KEYWORDS

ddRADseq, gene flow, geographical range, hybridisation, introgression, speciation

## INTRODUCTION

The role of interspecific gene flow in species diversification is now broadly recognised as a key evolutionary process, with outcomes ranging from introgression (i.e., the transfer of genes between species as a result of hybridisation and repeated backcrossing) to the formation of a new hybrid species, when admixed individuals acquire reproductive isolation from both parents (Abbott et al., 2013; Capblancq et al., 2015; Hedrick, 2013; Mallet, 2005; Nogueras & Ortego, 2022). The conditions under which hybridisation occurs are multifaceted, depending on the spatial distribution and degree of reproductive isolation amongst taxa. Hybridisation can occur when habitats are structured in mosaics, with parent species living in specific habitats and hybrids establishing in intermediate habitats or in habitats not used by parents (Larson et al., 2013; Sottas et al., 2018). This habitat mosaic promoting hybridisation is sometimes due to human activities that fragment habitats and create new contact zones (Bell & Irian, 2019). Pleistocene glacial cycles also promoted hybridisation by favouring contact between species that previously evolved in allopatry (Huck et al., 2012; Melo-Ferreira et al., 2007) during periods of glacier expansion or contraction. Similarly, current climate warming may drive species ranges to collide (Larson et al., 2019; Muhlfeld et al., 2014). Aside from geographic opportunity, reproductive isolation has to be incomplete for gene pools to mix and form hybrid individuals, which often requires that parent species be closely related (Price & Bouvier, 2002). Reproductive isolation is, furthermore, not constant across time because of climate change that can erode prezygotic ecological or phenological isolation (Vallejo-Marín & Hiscock, 2016), and across space, with different hybridisation rates between the same pair of species depending on the landscape (Larson et al., 2013).

The genus *Coenonympha* Hübner (Nymphalidae) is an ecologically diversified group of butterflies exhibiting a large diversity of range sizes, habitat preferences and climatic niches (Hausharter et al., 2023) amongst the ca. 32 currently recognised species (Bozano, 2002; Kodandaramaiah & Wahlberg, 2009). A recent phylogeny of the genus based on whole-genome sequences has highlighted extensive discordance across markers resulting from incomplete lineage sorting during rapid speciation and/or from post-speciation gene flow (Greenwood et al., 2025). Interspecific gene flow appears especially prevalent within the young and species-rich hero subclade. This group comprises 11 species which occupy lowland meadows or alpine grasslands of Europe, North Africa and Asia (Greenwood et al., 2025). While the *Coenonympha* genus presumably originated in the mountains of Central Asia, the hero group diversified predominantly in the West Palearctic (Kodandaramaiah & Wahlberg, 2009). In this species group, a

hybridisation event between the pearly heath *C. arcania* (Linné, 1761) and the alpine heath *C. gardetta* (Prunner, 1798) during the late Pleistocene produced a hybrid lineage that subsequently split in allopatry giving rise to Darwin's heath *C. darwiniana* (Staudinger, 1871) in Ticino (Switzerland) and *C. cephalidarwiniana* (Verity, 1953) in the Southern Alps (Capblancq et al., 2015) with limited gene flow persisting between the hybrid lineages and their parental species (Capblancq et al., 2019). *C. cephalidarwiniana* was named *C. macromma* in Capblancq et al., 2015, 2019, 2020, but holotype analysis shows that the correct name for this taxon is *C. cephalidarwinia* (Gallo & Bisi, 2020). The presence of hybrid individuals in contemporary contact zones points to incomplete reproductive isolation amongst these four species (Capblancq et al., 2020). Incomplete reproductive isolation is suspected to extend beyond the *C. gardetta*–*C. arcania* complex to other members of the hero group; indeed, strong phylogenetic discordance has been found in this group, including an uncertain placement of *C. leander* (Esper, 1784) as sister either to the boreal species *C. hero*, or to the alpine *gardetta*–*arcania* complex or even as part of this complex, depending on the genomic marker considered in mtDNA and nuclear DNA (Greenwood et al., 2025; Kodandaramaiah & Wahlberg, 2009).

The present study aims to clarify species boundaries and the extent of past and present gene flow amongst six species of the *Coenonympha* hero subclade that are distributed in continental Europe. Two are lowland butterflies with large and partly overlapping geographic distributions: the pearly heath *C. arcania* is found from Western Europe to the Urals, and the scarce heath *C. hero* (Linné, 1761) is distributed along a mid-to-high latitudinal band stretching from Europe to Japan. Two are highland species with a more restrained distribution: the alpine heath *C. gardetta* is found in Massif Central and throughout the Alps above 1500 m, and the Russian heath *C. leander* is distributed across mountain ranges spanning from the Balkans and the Caucasus to the Urals (Bozano, 2002). The last two are the hybrid species found in disjunct restricted highland areas in the Alps, *C. darwiniana* and *C. cephalidarwiniana*. The six species can be easily distinguished in the field by their wing pattern (eyespot position and colour, Bozano, 2002; Capblancq et al., 2015). It is notable that *C. arcania* is currently in geographical contact with all the other species in some part of its distribution, and all species are in contact with at least two other species. Our main objective is to determine whether the observed phylogenetic discordance can be attributed to past or current gene flow between these six species, or to hybrid speciation events beyond the already established case of the alpine hybrid species *C. darwiniana* and *C. cephalidarwiniana*. The hybrid origin of these two species was confidently assessed based on double-digest Restriction-site Associated DNA sequencing (ddRADseq) de

novo reconstructed loci, by comparing different speciation scenarios using approximate Bayesian Computation (ABC) using DIYABC (Capblancq et al., 2015). However, DIYABC does not allow for continuous gene flow. In the present study, we used a more flexible approach to compare speciation scenarios, including divergence with continuous gene flow (fastsimcoal2). Furthermore, a reference genome for *C. arcania* has been published (Legeai et al., 2024) and we mapped the ddRAD reads from all six species on this reference genome. We first evaluated the genetic structure of the whole dataset, and the genetic diversity within species. We then evaluated the extent of differentiation and admixture amongst species, and searched for signals of heterospecific gene flow. In order to determine whether hybrid speciation can be extended beyond the already described two hybrid species of the *Coenonympha* hero subclade, we focused on four species (*C. arcania*, *C. hero*, *C. gardetta*, *C. leander*) to compare speciation scenarios involving different levels of interspecific gene flow and hybrid speciation.

## MATERIALS AND METHODS

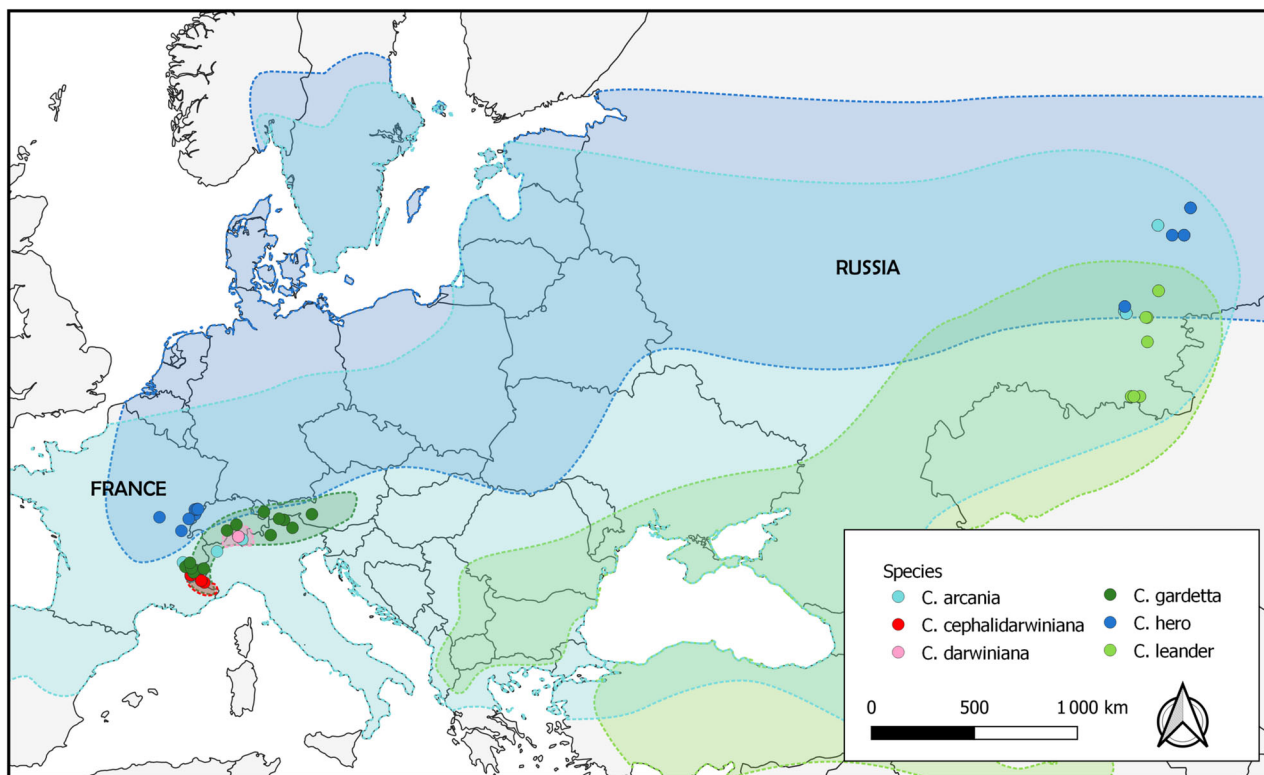
### Sampling

For each of the two lowland widespread species, two populations were sampled, in western Europe (*C. arcania* W and *C. hero* W) and at the southern extent of the Urals in Russia (*C. arcania* E and *C. hero* E).

The western population of *C. arcania* is composed of individuals from the Alps (France, Italy and Switzerland; Capblancq et al., 2015) and the western *C. hero* population comes from the Franche-Comté region of France which corresponds to the south-western margin of the species distribution (Sherpa et al., 2022). *C. gardetta* individuals were sampled from the whole range of the species across the Alps (France, Switzerland, Italy, Austria) and Massif Central. *Coenonympha darwiniana* was sampled in one locality in Switzerland, and *C. cephalidarwiniana* in the south of its distribution (southern French Alps) to avoid contact zones with *C. gardetta* where hybridisation still occurs and individuals are difficult to identify (Capblancq et al., 2019, 2020). *C. leander* individuals were sampled in six close localities in the Urals (Russia) (Figure 1). In total, 140 individuals were sampled, with a sample size from 14 to 26 individuals per population (Appendix S1). All specimens were sampled between 2013 and 2021. The four alpine species were sampled throughout the Alps for the study of Capblancq et al., 2015, *C. hero* from Jura was sampled for the study Sherpa et al., 2022, and all the populations from Russia were sampled in 2021 and genotyped for the present study.

### Genetic data acquisition

We used a subset of the ddRAD raw reads previously obtained for the *arcania-gardetta* species complex (Capblancq et al., 2015) and for *C. hero* (Sherpa et al., 2022) from western Europe and added



**FIGURE 1** Map showing the geographical distribution in Europe of each of the six studied *Coenonympha* species and the location of sampled butterflies. Each coloured point represents a sampled species/population (QGIS v3.28.10, QGIS.org, 2023).

25 *C. hero*, 26 *C. arcania*, and 17 *C. leander* newly genotyped individuals from Russia (Appendix S1). DNA was extracted from the thorax using the DNeasy Blood Tissue Kit (QIAGEN, Germany), and a double-digest Restriction-site Associated DNA (ddRAD) library was constructed using SbfI and MspI as described in (Sherpa et al., 2022), and paired-end sequenced on Novaseq Illumina (Novogene, UK).

Raw reads were demultiplexed using the process-radtags programme in Stacks v2.66 (Catchen et al., 2013) and mapped against the recently published *C. arcania* reference genome (Genbank ID: GCA\_036785405.1; Legeai et al., 2024) using the BWA-MEM algorithm implemented in BWA v0.7.17 software (Li & Durbin, 2009) with the -M argument to make sure reads mapped on two different portions are ignored by downstream callers. The Sequence Alignment Map (SAM) files obtained from BWA were compressed into Binary Alignment Map (BAM) files using the view command in SAMtools v1.13 (Danecek et al., 2021). The sort command was used to sort read alignments based on read position relative to the reference genome. Single nucleotide polymorphism (SNP) positions were retrieved with the Stacks pipeline using the sorted BAM files. Loci were identified in each individual using the gstacks programme with a minimum PHRED-scaled mapping quality of 20 (--min-mapq 20). The populations programme was then used for the first time to call SNPs on loci that were found in all populations (-p 8) at a minimum prevalence of 40% within each population (-r 0.4). To produce a SNP variant file without the known hybrids, Populations, was run a second time without *C. darwiniana* and *C. cephalidarwiniana* (with -p 6) but including monomorphic sites in the file (--vcf-all argument). Using VCFtools v0.1.16 (Danecek et al., 2011), variants mapped to the mitochondrial DNA and suspected W chromosome (scaffolds 31, 32, 33) were filtered out using the --not-chr argument. All sites with a mean read depth lower than 5 (--mean-minDP 5) were removed. More filters were applied with VCFtools to create 3 separate datasets for species differentiation ('Population' dataset), introgression ('Introgression' dataset) and speciation analyses ('Demographic' dataset). The 'Population' dataset for species differentiation analyses was produced by keeping only sites with a minimum minor allele count of 2 (--mac 2). Filtered data were thinned to retain a single SNP per 100 bp (--thin 100) to limit the impact of non-independence between variant sites. The 'Introgression' dataset was filtered to retain sites with a minimum minor allele count of 2 (--mac 2). The 'Demographic' dataset is a VCF file including all sites (polymorphic and monomorphic) genotyped for the 112 individuals from the four species *C. arcania*, *C. gardetta*, *C. leander* and *C. hero*. The two hybrid species (*C. darwiniana* and *C. cephalidarwiniana*) were not considered in the speciation scenarios to simplify models, and because their hybrid origin has already been established (Capblancq et al., 2015, 2019, 2020). A Site Frequency Spectrum (SFS) was created for each taxon as an input for downstream analyses using the R script vcf2sfs (Liu et al., 2018). As the ancestral allele was unknown, a folded SFS was created.

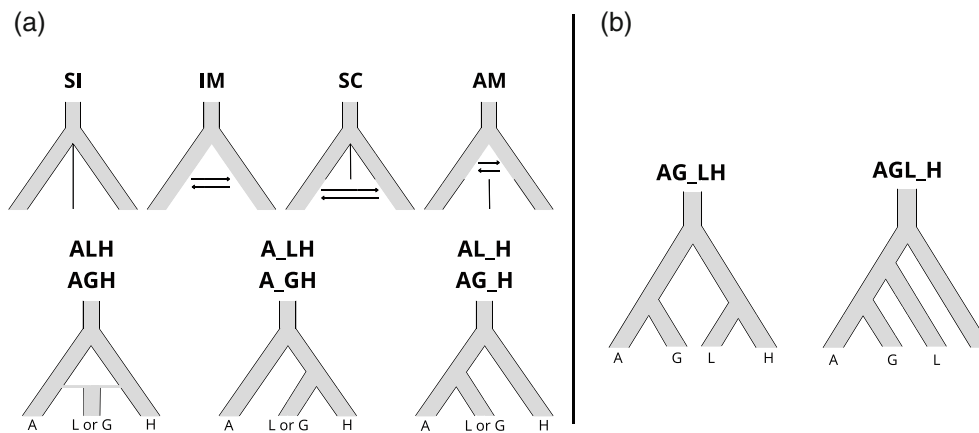
## Species differentiation

Different methods were used to evaluate the extent of differentiation amongst species. First, the pairwise fixation index ( $F_{st}$ ) of Weir and Cockerham (1984) was calculated to estimate the genetic distance

between each a priori population using the R-package hierfstat v0.5-11 (Goudet & Jombart, 2022). Second, sNMF (Frichot et al., 2014) was used to estimate the individual ancestry coefficients using the R-package LEA v3.14.0 (Frichot & François, 2015) with R v4.3.2 (R Core Team, 2023). This programme has the advantage of avoiding Hardy-Weinberg assumptions while producing similar results as STRUCTURE or ADMIXTURE in a shorter time (Frichot et al., 2014). A number of populations ( $K = 2-10$ ) were simulated for the 'Population' dataset, with 100 repetitions per  $K$ -value. The regularisation parameter ( $\alpha$ ), which penalises intermediate ancestry values, was set to 100. The best number of populations was chosen by looking at the value of  $K$  which produced the lowest cross-entropy criterion (CE) over all iterations, and the iteration with the lowest CE was chosen for plotting. Third, a principal component analysis (PCA) was performed on individuals to identify population clustering patterns using the R-package adegenet v2.1.10 (Jombart, 2008; Jombart & Ahmed, 2011). Finally, a maximum likelihood (ML) tree was performed based on the whole concatenated aligned sequences (56,515 sites present in all populations with less than 60% missing data) using IQtree v2.2.2.6 (Nguyen et al., 2015). The substitution model was selected by using ModelFinder (Kalyaanamoorthy et al., 2017) and 1000 UltrafastBootstraps (Hoang et al., 2018) were performed to build the ML tree.

## Detection of introgression

The ABBA-BABA test, also known as the Patterson's  $D$  statistic (Patterson et al., 2012), was used to detect introgression in subsets of species/populations while accounting for potential incomplete lineage sorting (ILS) in their history. The ABBA-BABA test is generally conducted on 4-population samples with the tree topology [((P1, P2), P3), O]], where O is a known outgroup, P1 and P2 are sister taxa, and P3 is a taxon which may have experienced past gene flow with either P1 or P2. In this framework, alleles present in the outgroup are set as ancestral (A), and the distribution of derived alleles (B) is examined in P2-P3 and P1-P3. Sites are expected to produce allelic patterns that match the phylogeny (i.e., BBAA) but the discordant patterns ABBA and BABA can also be observed. If this pattern is due to incomplete lineage sorting (ILS), that is random sorting of ancestral polymorphism between diverging sister species, ABBA and BABA, patterns should be observed at the same frequency. However, if gene flow has occurred after divergence, there should be a significant skew in the frequency of ABBA and BABA patterns. The  $D$  statistic appraises the difference in these site pattern frequencies. If significantly greater than 0, the statistic indicates an excess of ABBA patterns driven by gene flow between P2 and P3, while if it is lower than 0, it means that there is an excess of BABA patterns, highlighting gene flow between P1 and P3. The fraction of the genome involved in introgression, the  $f_4$ -statistic, can be derived by taking the ratio of excess site patterns to the total number of sites considered. These statistics were calculated for taxa in the 'Introgression' dataset using the Dsuite v0.5 r53 software (Malinsky et al., 2021). As the choice of an outgroup in this dataset proved challenging, the Dquartet command was used to



**FIGURE 2** Speciation scenarios simulated with fastsimcoal2. (a) Scenarios with two species (upper row) with arrows indicating gene flow between species and a vertical line indicating isolation. Scenarios with three species (bottom row) testing whether *C. leander* or *C. gardetta* are hybrid species between *C. arcania* and *C. hero* (ALH and AGH models), diverged from *C. hero* (A\_LH and A\_GH models) or from *C. arcania* (AL\_H and AG\_H models). Letters under each branch indicate the species (A = *C. arcania*, G = *C. gardetta*, L = *C. leander*, H = *C. hero*). (b) The two topologies simulated with *C. leander* diverging either from *C. hero* (AG\_LH model) or from *C. arcania* (AGL\_H model), each tested without (0) or with gene flow between *C. arcania* and *C. leander*, or/and *C. arcania* and *C. hero* (1,2).

perform the calculation of the D statistic and the  $f_4$ -statistic for every possible quartet of taxa with all potential outgroups. In the output of Dsuite, P1 and P2 are rearranged to ensure that the considered hybridisation event is always between P2 and P3, and that the test statistic is positive. This command was launched using 80 Jackknife blocks (-k 80) and the  $p$ -values were corrected using the False Discovery Rate (FDR) method according to the author's recommendations. An admixture tree was also created using the Treemix v1.13 software (Pickrell & Pritchard, 2012). This software tests for discrete gene flow events and directions by allowing poorly fitted branches to have multiple ancestry on the basis of a covariance matrix. Treemix was run with migration events from 0 to 8 (-m argument), and groups of 1, 100, 300, 500 and 1000 SNPs (-k argument). The most probable number of migration events amongst taxa was determined using the R-package OptM v0.1.6 (Fitak, 2021) with the Evanno method.

## Speciation scenarios

The 'Demographic' dataset was used to test different speciation scenarios using fastsimcoal2 v2.8 (Excoffier et al., 2013, 2021; Marchi et al., 2024). This software is a coalescent simulation programme in which parameters such as the time of divergence or the ancestral population size can be estimated by comparing simulated and observed SFS under a maximum likelihood framework. The estimation of parameters is done using an Expectation Conditional Maximisation algorithm (ECM) which starts with random values, then maximises each parameter in turn while keeping the others at the last estimated value to finally calculate the composite likelihood. To make the simulation simpler, and to place emphasis on interspecific demographic processes, western and eastern populations were merged for *C. hero* and *C. arcania*. We ran the fastsimcoal inferences using a wide parameter space to avoid any unwanted constraints. The following ranges

were used: N in [ $10^3$ - $10^6$ ] and m in [ $10^{-8}$ - $10^{-2}$ ]. The mutation rate was set to  $2.9e-9$  to match the rate estimated in another Nymphalid butterfly species, *Heliconius melpomene* (Keightley et al., 2015). As the considered *Coenonympha* species are univoltine, the number of generations per year was set to 1.

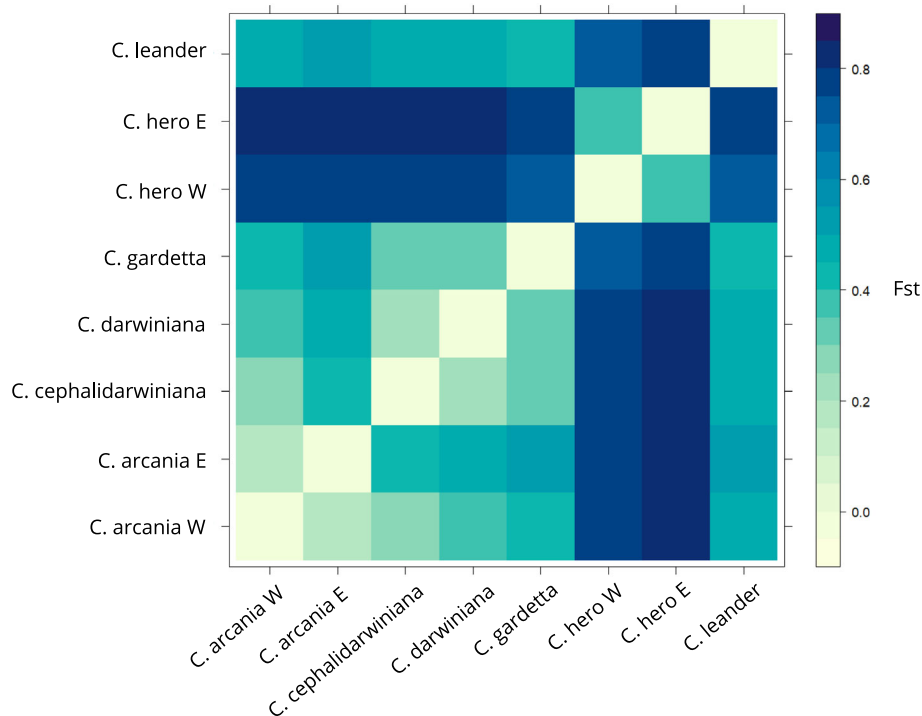
Before designing speciation scenarios involving the 4 species, we tested all the possible two-species scenarios. Four scenarios (Figure 2a) were simulated for each pair of species: Strict Isolation (SI), Isolation with Migration (IM), Ancient Migration (AM) and Secondary Contact (SC). The SI scenario implies that no gene flow occurred after species divergence, the IM scenario implies that species have diverged while experiencing continuous gene flow, the AM scenario implies that gene flow occurred initially during divergence but ceased at some point in the past, and the SC scenario implies that species diverged in isolation initially but that gene flow resumed later and may still occur now. We then tested scenarios with three populations to evaluate whether *C. gardetta* and *C. leander* diverged from *C. arcania*, *C. hero*, or are the result of an admixture event between those two (hybrid speciation; Figure 2a). Finally, we designed 12 four-species scenarios including two alternative strictly bifurcating tree topologies (Figure 2b) with different levels of gene flow and three scenarios involving hybrid speciation (Appendix S3). As gene flow is known to occur between *C. arcania* and *C. gardetta* (Capblancq et al., 2019), migration was simulated between these two species in both directions in all scenarios.

For each scenario, parameters were estimated using 50 runs of 100 ECM cycles (-L 100) of 500,000 simulations (-n 500,000) with a random seed for each run. By looking at the lowest difference between the observed log-likelihood and the estimated log-likelihood, the best run was chosen for each scenario. Then, the best scenario was selected by calculating the Akaike Information Criterion (AIC), which estimates the quality of a model while taking into account the number of parameters to prevent overfitting. The run with the lowest

**TABLE 1** Genetic diversity indices for each of the studied populations, calculated from the 'Demographic dataset' (525 ddRAD loci; 56,679 sites, 8861 SNPs).

Population	Sample size	Sites	Variant Sites	Polymorphic Sites	% polymorphic sites	Ho	He	$\pi$
<i>C. arcania</i> W	15	56,679	8861	2745	4.843	0.007	0.011	0.012
<i>C. arcania</i> E	26	56,679	8861	2440	4.305	0.006	0.010	0.010
<i>C. cephalidarwiniana</i>	14	56,679	8861	2642	4.661	0.007	0.011	0.012
<i>C. darwiniana</i>	14	56,679	8861	2451	4.324	0.007	0.011	0.012
<i>C. gardetta</i>	15	56,679	8861	2873	5.069	0.006	0.012	0.013
<i>C. hero</i> W	14	56,679	8861	342	0.603	0.002	0.002	0.002
<i>C. hero</i> E	25	56,679	8861	554	0.977	0.002	0.002	0.002
<i>C. leander</i>	17	56,679	8861	2761	4.871	0.006	0.012	0.012

Abbreviations: He, expected heterozygosity; Ho, observed heterozygosity;  $\pi$ , nucleotide diversity.

**FIGURE 3** Pairwise Weir and Cockerham  $F_{st}$  between each population. Darker colours indicate higher values of  $F_{st}$  and thus higher differentiation between populations.

AIC was retained per scenario, and the  $\Delta$ AIC was calculated between scenarios, using  $\Delta$ AIC values higher than 10 as evidence for significant improvement of one model over another. All files used to simulate scenarios under fastsimcoal2 are available on GitHub: <https://github.com/Camizuli/Coenonympha-FSC2>.

## RESULTS

### Genetic diversity within and between populations and species

All 140 individuals mapped well against the reference genome, with <2% unmapped sites, and a mean depth coverage of 41.7

(Appendix S2). After filtering, 525 loci remained with 8861 SNPs in the dataset including the two hybrid species, and 609 loci remained with 8695 SNPs and 56,595 monomorphic sites in the dataset without the two hybrid species ( $n = 112$ ). The 'Population' dataset included 476 independent SNPs, and the 'Introgression' and the 'Demographic' datasets contained, respectively, 7311 and 8659 SNPs.

The genetic diversity indices were very similar in all the populations, ranging from 4% to 5% polymorphic sites, and from 0.010 to 0.013 for the nucleotide diversity  $\pi$ , except for the two populations of *C. hero*, which were much less diverse (Table 1). Pairwise  $F_{st}$  values (Figure 3) show that *C. hero* is strongly differentiated from the other species with extremely high values (>0.7) and is also strongly structured between the eastern and western populations ( $F_{st} \approx 0.36$ ). In

contrast, *C. arcania* eastern and western populations are much less differentiated ( $F_{st} \approx 0.17$ ) and exhibit some genetic proximity with *C. cephalidarwiniana* ( $F_{st} \approx 0.29$ ) and *C. darwiniana* ( $F_{st} \approx 0.37$ ). *C. gardetta* is equally divergent from *C. cephalidarwiniana* and *C. darwiniana* (respectively  $F_{st} \approx 0.32$  and  $0.35$ ), as expected between hybrid lineages and their parents. It is notable that *C. gardetta* seems equally close to *C. leander* and to the western population of *C. arcania* ( $F_{st} \approx 0.42$ ) but more distant from the eastern population of *C. arcania* ( $F_{st} \approx 0.50$ ). *C. leander* has almost the same  $F_{st}$  value with both hybrid species ( $F_{st} \approx 0.45$ ) which is slightly higher than the value with *C. gardetta*. The two hybrid species have very similar  $F_{st}$  values, but *C. cephalidarwiniana* has a lower value with western *C. arcania* (Figure 3).

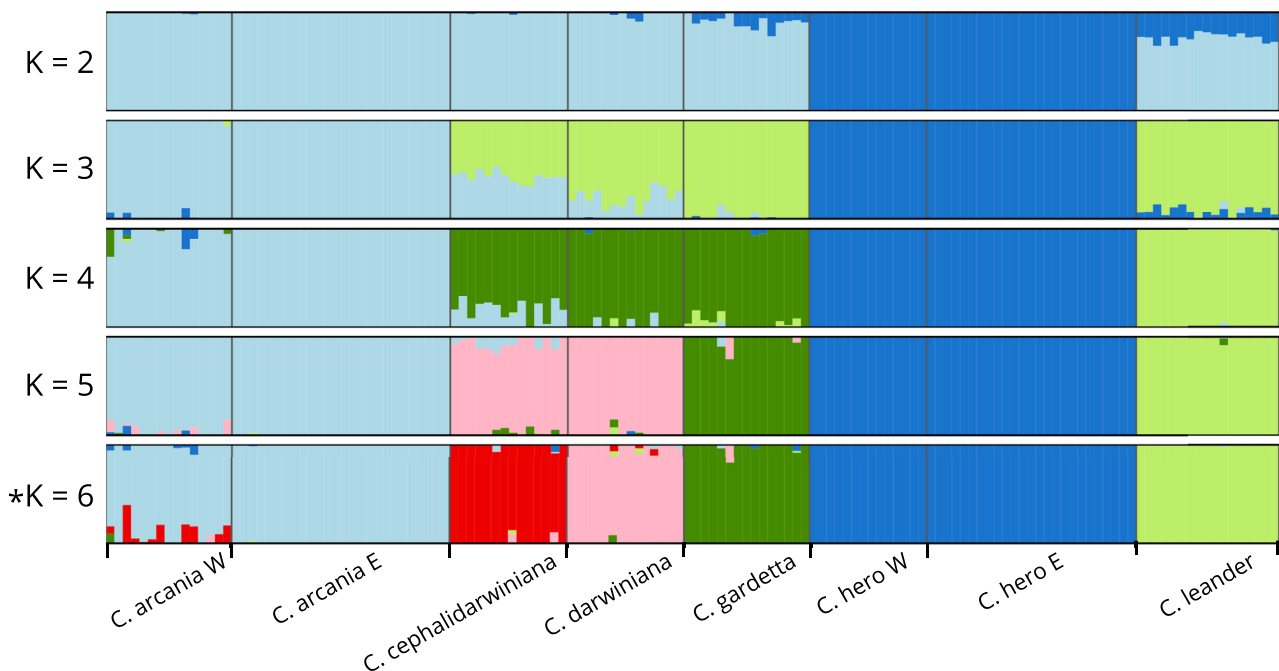
The sNMF clustering analysis indicates that the best supported number of populations is 6, the number of species present (Figure 4). At  $K = 2$ , the first split is between *C. hero* and all other species, but *C. leander*, and to a lesser extent *C. gardetta*, have some shared ancestry with the *C. hero* cluster. At  $K = 3$ , the *C. arcania* cluster is partly separated from other clusters. *C. leander* still shares some ancestry with *C. hero*, which is no longer the case for *C. gardetta*. *C. cephalidarwiniana* is half of the *C. arcania* cluster and half of the *C. gardetta*-*C. leander* cluster, while *C. darwiniana* shares less ancestry with *C. arcania*. At  $K = 4$ , the *C. gardetta*-*C. leander* cluster is split into an almost pure *C. leander* cluster and a cluster including *C. gardetta* with the two hybrid species, but some *C. gardetta* individuals share an ancestry with *C. leander*. *C. cephalidarwiniana* still shares some ancestry with *C. arcania*, and some western *C. arcania* individuals share very small amounts of ancestry with the *C. hero* and *C. gardetta* clusters. At  $K = 5$ , *C. gardetta* is separated from both hybrid species. Some

individuals from the western population of *C. arcania* still have a part of the *C. hero* cluster and from the hybrid cluster. Finally at  $K = 6$ , each species forms its own cluster, with pure clusters for *C. hero* and *C. leander*, while the remaining species show weak signals of admixture. Some individuals of *C. darwiniana* have a small part of ancestry from *C. cephalidarwiniana* and vice versa, but also from *C. leander* and *C. arcania*. One *C. gardetta* individual has a large amount of *C. darwiniana* ancestry ( $\approx 20\%$ ), and some others have tracks of ancestry with *C. hero* and *C. arcania*. The eastern population of *C. arcania* is pure but the western population shares a large amount of ancestry with *C. cephalidarwiniana* (up to 38%), and a small amount with *C. hero*.

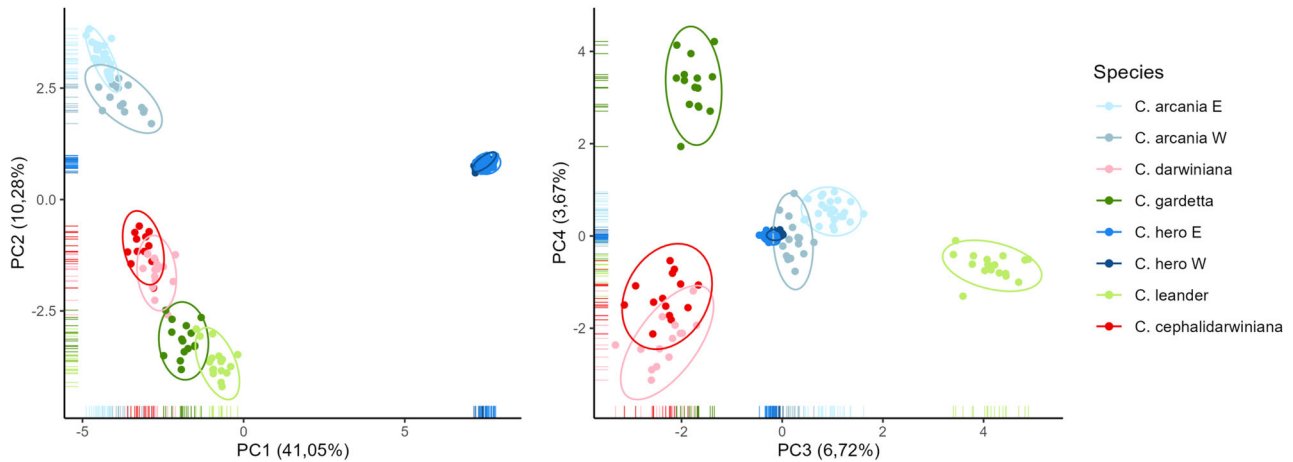
The principal component analysis (PCA) shows similar sequential species differentiation along four axes that capture more than 50% of the total genetic variation (Figure 5). PC1 separates all taxa from *C. hero*. The two hybrid species strongly overlap and fall between their parent species (*C. arcania* and *C. gardetta*), as expected for hybrid taxa. PC2 separates *C. arcania* from the other taxa, PC3 separates *C. leander* from *C. gardetta*, while the two hybrid species *C. cephalidarwiniana* and *C. darwiniana*, that still partly overlap, are separated from *C. gardetta* on PC4. The ML tree shows that the six species are reciprocally monophyletic with high bootstrap support (Appendix S4).

## Detection of introgression

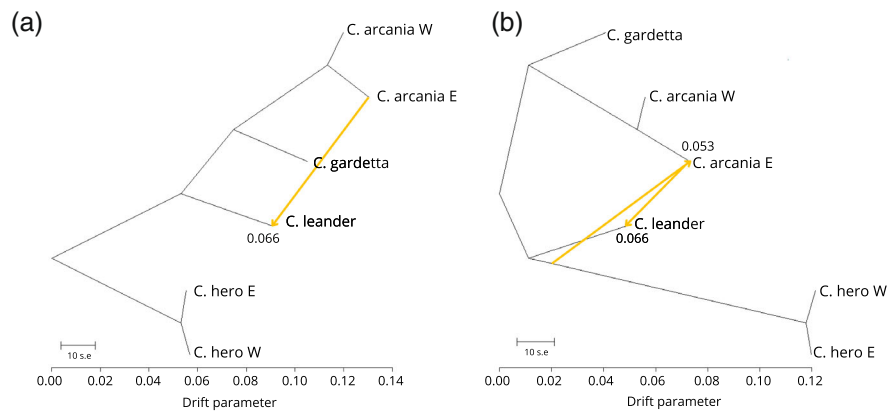
The ABBA-BABA tests and the Treemix analysis performed with *C. cephalidarwiniana* and *C. darwiniana* recovered gene flow from



**FIGURE 4** Population genetic structure of the 140 individuals with number of clusters from  $K = 2$  to 6. Each colour represents a genetic cluster, and barplots the admixture proportion for each individual. The star (\*) indicates the best number of clusters according to the cross-entropy criterion.



**FIGURE 5** Principal component analysis of the four first axes that altogether represent more than 50% of total genetic variation. Each point represents an individual and is coloured according to the species it belongs to. Ellipses regroup 95% of points for each population.



**FIGURE 6** Relationships and migration edges inferred by Treemix with (a) 1 migration and (b) 2 migrations events without *C. cephalidarwiniana* and *C. darwiniana*. Migration edges are coloured according to the migration weight, and labels at the tip of the arrows indicate the exact value. The scale bar indicates 10 times the average standard error. The drift parameter is the amount of genetic drift within each population.

*C. gardetta* to the ancestor of both hybrids (Appendix S5). As these migration events are known and recent, they may hide other events that are older. In order to detect other putative events, the same analysis was done without *C. cephalidarwiniana* and *C. darwiniana*, using the 'Introgression 2' dataset (Appendix S2). Applying Treemix to this dataset, the best number of migration events is one, in a scenario where *C. arcania* E and *C. leander* share 6% of their alleles due to gene flow (Figure 6a). When allowing two migration edges, an additional migration event links the ancestor of *C. hero* to *C. arcania* E, with 5% of alleles shared (Figure 6b). The ABBA-BABA test recovered introgression between *C. arcania* W and *C. gardetta* with a  $f_4$ -statistic between 0.12 and 0.14. Moreover, introgression was found between *C. leander* and both populations of *C. arcania*, in particular with the eastern population. With the two *C. arcania* populations in P1 and P2, no introgression was detected with *C. leander*, while introgression from *C. leander* was detected when the western and eastern populations of *C. arcania* were merged, suggesting that gene flow predates the splitting of these populations (Table 2).

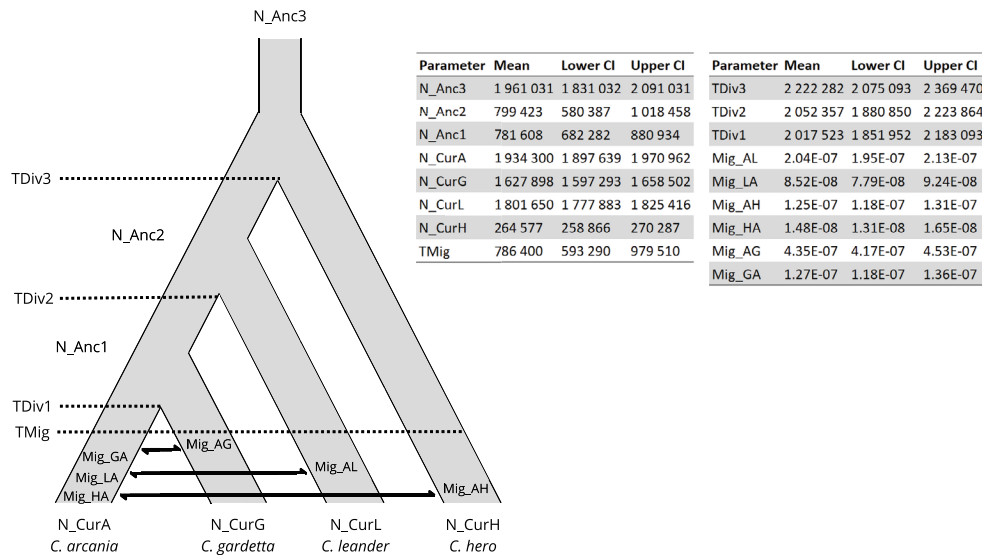
## Pervasive gene flow during the diversification of the four species

The strict isolation (SI) scenario was rejected for all pairs of species, indicating the presence of gene flow accompanying the diversification of the four species. However, the  $\Delta AIC$  was less than 10 between the other scenarios (IM, AM and SC) for all species pairs, meaning that these scenarios were equally good, except for *C. gardetta* and *C. leander* for which AM was the best scenario (Appendix S6). With three species, the best scenario was hybrid speciation (ALH model), with *C. leander* resulting from an admixture between *C. arcania* and *C. hero*, each respectively contributing 68% and 32%. For the triplet with *C. gardetta*, the hybrid speciation scenario (AGH model) and divergence from *C. arcania* scenario (AG\_H model) were equivalent ( $\Delta AIC < 10$ ). Based on these results, it appeared that both *C. leander* and *C. gardetta* might be hybrid species, a notion also suggested by sNMF and Treemix results. This hypothesis was explored by simulating an additional 12 divergence scenarios considering the four species

**TABLE 2** Results of the ABBA BABA tests for every population quartets.

P1	P2	P3	P4	Dstatistic	f <sub>4</sub> -statistic	BBAA	ABBA	BABA	p-value
<i>C. arcania E</i>	<i>C. arcania W</i>	<i>C. gardetta</i>	<i>C. leander</i>	0.112	0.119	332.394	87.415	69.875	<0.001
<i>C. arcania E</i>	<i>C. arcania W</i>	<i>C. gardetta</i>	<i>C. hero</i>	0.078	0.142	352.168	82.773	70.774	0.041
<i>C. arcania W</i>	<i>C. arcania E</i>	<i>C. leander</i>	<i>C. hero</i>	0.053	0.089	446.514	55.206	49.666	0.215
<i>C. gardetta</i>	<i>C. arcania W</i>	<i>C. leander</i>	<i>C. hero</i>	0.096	0.176	235.081	107.628	88.752	0.041
<i>C. gardetta</i>	<i>C. arcania E</i>	<i>C. leander</i>	<i>C. hero</i>	0.119	0.19	225.134	115.22	90.804	0.022
<i>C. gardetta</i>	<i>C. arcania</i>	<i>C. leander</i>	<i>C. hero</i>	0.112	0.082	228.855	112.621	89.965	0.012

Note: The p-values are corrected using the FDR method. In the last line *C. arcania* = *C. arcania E* + *W*.



**FIGURE 7** Best speciation scenario between the 4 *Coenonympha* species. Parameters are the ancestral (N\_Anc1, N\_Anc2, N\_Anc3) and current effective population sizes (N\_CurA, N\_CurG, N\_CurL, N\_CurH) in haploid number, times of divergence (TDiv1, TDiv2, TDiv3), and migration rates (Mig\_AL, Mig\_LA, Mig\_AH, Mig\_HA). Table shows the mean and 95% confidence interval for each parameter estimate from the 10 best runs (out of 50) with fastsimcoal2.

(Appendices S3 and S7). Three of these scenarios involved hybrid speciation with *C. arcania* and *C. hero* as parental species: *C. leander* is a hybrid species (AGLH1), both *C. leander* and *C. gardetta* originated from two distinct hybridisation events (AGLH2), and finally a cascade of hybridisation was hypothesised, with a first hybridisation event between *C. arcania* and *C. hero* at the origin of *C. leander* followed by a second hybridisation event between *C. leander* and *C. arcania* giving rise to *C. gardetta* (AGLH2nest; Appendix S6). The other scenarios (strictly bifurcating, with or without gene flow, gene flow continuous, or secondary contact) were designed based on the migration edges identified in the Treemix results. The best model (AGL\_HSC2) did not involve hybrid speciation and is concordant with the Treemix results: *C. leander* diverged from *C. arcania* and two migration events (Secondary Contact) link *C. leander* with *C. arcania* and *C. hero* with *C. arcania*, respectively. Divergence time between *C. arcania* and *C. gardetta* was estimated at 2.01 Mya, at 2.05 Mya between *C. leander* and *C. arcania*, and 2.22 Mya between *C. hero* and *C. arcania*. Secondary contact took place around 790kya, with slightly asymmetrical migration rates from *C. arcania* to *C. gardetta* (4.35e-7 vs. 1.27e-7), from *C. arcania* to *C. leander* (2.04e-7 vs. 8.52e-8), and from *C. arcania* to *C. hero* (1.25e-7 vs. 1.48e-8) (Figure 7).

## DISCUSSION

Our ddRAD analysis confirms that the six species under study are reciprocally monophyletic, consistent with their morphological distinctions and current taxonomic classification. However, the results also reveal extensive historical gene flow throughout their diversification. Notably, we confirm the hybrid origin of *C. cephalidarwiniana* and *C. darwiniana*, and demonstrate that hybridisation extends beyond this alpine species complex. It involves phylogenetically distant lineages that diverged prior to the Pleistocene glacial cycles (Greenwood et al., 2025) but have come into secondary contact since the late Pleistocene.

### A complex history of rapid divergence and secondary gene flow

Both the ABBA-BABA test and Treemix analysis recovered the known history of gene flow between *C. gardetta* and *C. arcania* (Capblancq et al., 2015, 2019, 2024), and detected introgression between *C. leander* and *C. arcania* primarily involving the eastern population of *C. arcania*, which co-occurs with *C. leander* in the Urals where our

sampling took place. Furthermore, Treemix identified gene flow between *C. arcania* E and *C. hero*, again between populations in close geographical proximity. In both cases, it is important to note that our sampling of these largely distributed species was limited, and extinct or unsampled populations that were involved in hybridisation with any species of this complex could have influenced the detected gene flow (i.e., ‘ghost’ introgression) (Tricou et al., 2022) preventing us from elaborating on the geographical context of hybridisation.

In accordance with the introgression tests, the two-species simulations with every possible pair firmly rejected the ‘Strict Isolation’ scenario but could not distinguish a best scenario amongst those that involved migration, except for *C. leander* and *C. gardetta* where an ‘Ancient Migration’ scenario is favoured (Appendix S5). These two species are the only currently allopatric species; all the other pairs of species involving either *C. arcania* or *C. hero* have some degree of geographical overlap. The three-species simulations favoured a hybrid speciation scenario for *C. leander* arising from admixture between *C. arcania* and *C. hero* but were not conclusive for *C. gardetta*. These three-species hybrid speciation scenarios were simplified, as they did not allow for additional migration to take place after the hybridisation event. For the four-species scenarios, we systematically allowed gene flow between *C. arcania* and *C. gardetta* (Capblancq et al., 2019) and compared strictly bifurcating scenarios with different levels of gene flow (ancient or recent) to scenarios involving one or two hybridisation events promoting the rise of *C. leander* and/or *C. gardetta* (hybrid speciation scenarios). However, these hybrid speciation scenarios were not supported. The best four-species scenario groups *C. leander* with *C. arcania* and *C. gardetta*, which contrasts original hypotheses that *C. leander* and *C. hero* are sister species (Kodandaramaiah & Wahlberg, 2009) and a more recent species tree recovered using whole genome sequences (Greenwood et al., 2025), but our present topology is reflected by the mitogenome and by part of the nuclear genome (Greenwood et al., 2025). The time estimate for the *C. arcania*–*C. gardetta* split (around 2 Mya) fits the divergence time previously estimated for these species (Capblancq et al., 2015) as well as the migration rate (around  $10^{-7}$ ) and effective population sizes (larger for *C. arcania* as compared with *C. gardetta*). The split of *C. leander* and *C. hero* is not much older, suggesting that these four species diverged in a short time period at the onset of the Pleistocene, which is also suggested by poorly resolved phylogenies and cyto-nuclear discordances (Greenwood et al., 2025). The best scenario involved two secondary gene flow events between *C. leander* and *C. arcania*, and between *C. hero* and *C. arcania*, at approximately the same time, around 800,000 years ago. This corresponds to a time of increased amplitude of the Pleistocene glacial cycles (more extreme temperatures, lasting longer) which might have more profoundly affected the species ranges than previous cycles, putting into contact species that previously evolved in allopatry. One limitation of these demographic inferences is that we might miss unknown/unsampled/extinct lineages that may have been involved in gene flow. These ‘ghost’ populations can have a large influence on demographic inferences (Yamahira et al., 2023). A possible ‘ghost’ population could be the enigmatic Balkanic heath *C. orientalis* (Rebel, 1913) which flies in the Balkans and does not have a clear taxonomic position, with some

authors arguing that it is a sub-species of *C. gardetta* or *C. leander* or that it falls within the *C. gardetta*–*C. arcania* complex (Coutsis & Ghalavás, 2005).

Our study indicates that gene flow extends beyond the alpine species complex to include the mountain species *C. leander* and the boreal species *C. hero*, both of which exchange genetic material with *C. arcania*. These three species currently coexist broadly in the Urals, where our sampling was conducted. Further investigation of populations in intermediate regions, such as the Balkans, could provide deeper insights into their gene flow history. *C. arcania* appears central to the evolutionary dynamics of this ecologically diverse group of European butterflies, exchanging genes with all studied species. It would be particularly valuable to determine whether *C. arcania* has transferred or acquired adaptive traits from other species, as observed in the two hybrid species (Doniol-Valcroze et al., 2024, 2025; Nève & Després, 2020).

## AUTHOR CONTRIBUTIONS

**Thibaud Camizuli:** Writing – original draft; formal analysis; data curation; methodology; investigation; writing – review and editing. **Matthew Greenwood:** Conceptualization; investigation; supervision; writing – review and editing. **Delphine Rioux:** Investigation; writing – review and editing. **Elena Zakharova:** Resources. **Laurence Després:** Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; supervision; project administration.

## ACKNOWLEDGEMENTS

We thank Thibaut Capblancq and Vincent Lefort for helpful discussions and comments on the manuscript.

## FUNDING INFORMATION

This work has been supported by grants from the French National Research Agency (Grant No. ANR-20-CE02-0017), and from the French Ministry of Education and Research. The sample collection in Russia was supported by the state assignment of the Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences (No. 122021000091-2). All the computations were performed using the GRICAD infrastructure (<https://gricad.univ-grenoble-alpes.fr>), which is supported by Université Grenoble Alpes.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The raw sequences (fastq) analysed in the present study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB85980 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB85980>).

## ORCID

Thibaud Camizuli  <https://orcid.org/0009-0000-0965-6805>

Matthew Greenwood  <https://orcid.org/0000-0002-0561-8868>

Elena Zakharova  <https://orcid.org/0000-0002-6704-4492>

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

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N. et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246. Available from: <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Bell, R.C. & Irian, C.G. (2019) Phenotypic and genetic divergence in reed frogs across a mosaic hybrid zone on São Tomé Island. *Biological Journal of the Linnean Society*, 128, 672–680. Available from: <https://doi.org/10.1093/biolinnean/blz131>
- Bozano, G.C. (2002) *Guide to the Butterflies of the Palearctic Region. Satyrinae, part III. Tribe Satyrini, subtribes Melanargiina and Coenonymphina, Melanargia, Coenonympha, Sinonympha, Triphysa*. Milano: Omnes Partes.
- Capblancq, T., Després, L. & Mavárez, J. (2020) Genetic, morphological and ecological variation across a sharp hybrid zone between two alpine butterfly species. *Evolutionary Applications*, 13, 1435–1450. Available from: <https://doi.org/10.1111/eva.12925>
- Capblancq, T., Després, L., Rioux, D. & Mavárez, J. (2015) Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, 24, 6209–6222. Available from: <https://doi.org/10.1111/mec.13479>
- 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. Available from: <https://doi.org/10.1002/ece3.5220>
- Capblancq, T., Boyer, F., Legeai, F., Joron, M. & Després, L. (2024) Untangling the contribution of adaptive versus non-adaptive processes in the evolution of reproductive isolation between *Coenonympha* butterflies. *bioRxiv*, 2024–11. <https://doi.org/10.1101/2024.11.29.625973>.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. Available from: <https://doi.org/10.1111/mec.12354>
- Coutsis, J.G. & Ghalvalás, N. (2005) A recently discovered new locality for *Coenonympha leander* in Greece, and notes about the taxonomic position of the species-group taxon *Coenonympha orientalis* (Lepidoptera: Nymphalidae, Satyrinae). *Phegea*, 33(4), 121–128.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158. Available from: <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O. et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, 10, giab008. Available from: <https://doi.org/10.1093/gigascience/giab008>
- Doniol-Valcroze, P., Després, L. & Joron, M. (2024) Strong differences in egg heat tolerance between closely related lowland and alpine butterfly species. *Ecological Entomology*, 6, 598–603. Available from: <https://doi.org/10.1111/een.13323>
- Doniol-Valcroze, P., Develay Nguyen, L., Buatois, B., Dötterl, S., Fuchs, R., Després, L. et al. (2025) Non-random sorting of parental chemical compounds during hybrid speciation. *Journal of Evolutionary Biology*, 38(5), 559–571. Available from: <https://doi.org/10.1093/jeb/voaf022>
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. & Foll, M. (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905. Available from: <https://doi.org/10.1371/journal.pgen.1003905>
- Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A. & Sousa, V.C. (2021) fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*, 37, 4882–4885. Available from: <https://doi.org/10.1093/bioinformatics/btab468>
- Fitak, R.R. (2021) OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, 6, bpab017. Available from: <https://doi.org/10.1093/biomethods/bpab017>
- Frichot, E. & François, O. (2015) LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925–929. Available from: <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. (2014) Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196, 973–983. Available from: <https://doi.org/10.1534/genetics.113.160572>
- Gallo, E. & Bisi, A. (2020) Considerazioni tassonomiche su *Coenonympha arcania* macromma Turati & Verity 1911 ed altri taxa ad essa correlati. *Annali del Museo Civico di Storia Naturale "G. Doria"*, Genova, 112, 325–334.
- Goudet, J. & Jombart, T. (2022) hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.5–11. <https://github.com/jgx65/hierfstat>
- Greenwood, M.P., Capblancq, T., Wahlberg, N. & Després, L. (2025) Whole genome data confirm pervasive gene discordance in the evolutionary history of *Coenonympha* (Nymphalidae) butterflies. *Molecular Phylogenetics and Evolution*, 202, 108222. Available from: <https://doi.org/10.1016/j.ympev.2024.108222>
- Haussharter, J., Rashid, S., Wessely, J., Strutzenberger, P., Moser, D., Gattringer, A. et al. (2023) Niche breadth explains the range size of European-centred butterflies, but dispersal ability does not. *Global Ecology and Biogeography*, 32, 1535–1548. Available from: <https://doi.org/10.1111/geb.13717>
- Hedrick, P.W. (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22, 4606–4618. Available from: <https://doi.org/10.1111/mec.12415>
- Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. Available from: <https://doi.org/10.1093/molbev/msx281>
- Huck, S., Büdel, B. & Schmitt, T. (2012) Ice-age isolation, postglacial hybridization and recent population bottlenecks shape the genetic structure of *Meum athamanticum* in Central Europe. *Flora - Morphology, Distribution, Functional Ecology of Plants*, 207, 399–407. Available from: <https://doi.org/10.1016/j.flora.2012.03.005>
- Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405. Available from: <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070–3071. Available from: <https://doi.org/10.1093/bioinformatics/btr521>
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermini, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589. Available from: <https://doi.org/10.1038/nmeth.4285>
- Keightley, P.D., Pinharanda, A., Ness, R.W., Simpson, F., Dasmahapatra, K.K., Mallet, J. et al. (2015) Estimation of the Spontaneous Mutation Rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, 32, 239–243. Available from: <https://doi.org/10.1093/molbev/msu302>
- Kodandaramaiah, U. & Wahlberg, N. (2009) Phylogeny and biogeography of *Coenonympha* butterflies (Nymphalidae: Satyrinae) – patterns of colonization in the Holarctic. *Systematic Entomology*, 34, 315–323. Available from: <https://doi.org/10.1111/j.1365-3113.2008.00453.x>
- Larson, E.L., Guilherme Becker, C., Bondra, E.R. & Harrison, R.G. (2013) Structure of a mosaic hybrid zone between the field crickets *Gryllus firmus* and *G. pennsylvanicus*. *Ecology and Evolution*, 3, 985–1002. Available from: <https://doi.org/10.1002/ece3.514>
- Larson, E.L., Tinghitella, R.M. & Taylor, S.A. (2019) Insect hybridization and climate change. *Frontiers in Ecology and Evolution*, 7, 348. Available from: <https://doi.org/10.3389/fevo.2019.00348>
- Legeai, F., Romain, S., Capblancq, T., Doniol-Valcroze, P., Joron, M., Lemaître, C. et al. (2024) Chromosome-level assembly and

- annotation of the pearly heath *Coenonympha arcania* butterfly genome. *Genome Biology and Evolution*, 16, evae055. Available from: <https://doi.org/10.1093/gbe/evae055>
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760. Available from: <https://doi.org/10.1093/bioinformatics/btp324>
- Liu, S., Ferchaud, A.-L., Grønkvær, P., Nygaard, R. & Hansen, M.M. (2018) Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Molecular Ecology*, 27, 4725–4743. Available from: <https://doi.org/10.1111/mec.14782>
- Malinsky, M., Matschiner, M. & Svardal, H. (2021) Dsuite – fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21, 584–595. Available from: <https://doi.org/10.1111/1755-0998.13265>
- Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20, 229–237. Available from: <https://doi.org/10.1016/j.tree.2005.02.010>
- Marchi, N., Kapopoulou, A. & Excoffier, L. (2024) Demogenomic inference from spatially and temporally heterogeneous samples. *Molecular Ecology Resources*, 24, e13877. Available from: <https://doi.org/10.1111/1755-0998.13877>
- Melo-Ferreira, J., Boursot, P., Randi, E., Kryukov, A., Suchentrunk, F., Ferrand, N. et al. (2007) The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. *Molecular Ecology*, 16, 605–618. Available from: <https://doi.org/10.1111/j.1365-294X.2006.03166.x>
- Muhlfeld, C.C., Kovach, R.P., Jones, L.A., Al-Chokhachy, R., Boyer, M.C., Leary, R.F. et al. (2014) Invasive hybridization in a threatened species is accelerated by climate change. *Nature Climate Change*, 4, 620–624. Available from: <https://doi.org/10.1038/nclimate2252>
- Nève, G. & Després, L. (2020) Cold adaptation across the elevation gradient in an alpine butterfly species complex. *Ecological Entomology*, 45, 997–1003. Available from: <https://doi.org/10.1111/een.12875>
- Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. Available from: <https://doi.org/10.1093/molbev/msu300>
- Noguerales, V. & Ortego, J. (2022) Genomic evidence of speciation by fusion in a recent radiation of grasshoppers. *Evolution*, 76, 2618–2633. Available from: <https://doi.org/10.1111/evo.14508>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y. et al. (2012) Ancient admixture in human history. *Genetics*, 192, 1065–1093. Available from: <https://doi.org/10.1534/genetics.112.145037>
- Pickrell, J. & Pritchard, J. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, 8(11), e1002967. Available from: <https://doi.org/10.1038/npre.2012.6956.1>
- Price, T.D. & Bouvier, M.M. (2002) The evolution of F1 postzygotic incompatibilities in birds. *Evolution*, 56, 2083–2089. Available from: <https://doi.org/10.1111/j.0014-3820.2002.tb00133.x>
- QGIS.org. (2023) QGIS geographic information system.
- R Core Team. (2023) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Sherpa, S., Kebaïli, C., Rioux, D., Guéguen, M., Renaud, J. & Després, L. (2022) Population decline at distribution margins: Assessing extinction risk in the last glacial relictual but still functional metapopulation of a European butterfly. *Diversity and Distributions*, 28, 271–290. Available from: <https://doi.org/10.1111/ddi.13460>
- Sottas, C., Reif, J., Kuczyński, L. & Reifová, R. (2018) Interspecific competition promotes habitat and morphological divergence in a secondary contact zone between two hybridizing songbirds. *Journal of Evolutionary Biology*, 31, 914–923. Available from: <https://doi.org/10.1111/jeb.13275>
- Tricou, T., Tannier, E. & de Vienne, D.M. (2022) Ghost lineages highly influence the interpretation of introgression tests. *Systematic Biology*, 71, 1147–1158. Available from: <https://doi.org/10.1093/sysbio/syac011>
- Vallejo-Marín, M. & Hiscock, S.J. (2016) Hybridization and hybrid speciation under global change. *New Phytologist*, 211, 1170–1187. Available from: <https://doi.org/10.1111/nph.14004>
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370. Available from: <https://doi.org/10.2307/2408641>
- Yamahira, K., Kobayashi, H., Kakioka, R., Montenegro, J., Masengi, K.W.A., Okuda, N. et al. (2023) Ghost introgression in ricefishes of the genus *Adrianichthys* in an ancient Wallacean lake. *Journal of Evolutionary Biology*, 36, 1484–1493. Available from: <https://doi.org/10.1111/jeb.14223>

## SUPPORTING INFORMATION

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

**Appendix S1.** List of the 140 individuals analysed, with geographical information and European Nucleotide Archive (ENA) repository (<http://www.ebi.ac.uk/ena>) accession number (fastq reads).

**Appendix S2.** Genomic statistics for each species/population for each dataset.

**Appendix S3.** The four-species scenarios tested involving simple divergence (‘Strictly bifurcating’) or hybrid speciation and different levels of gene flow, their number of parameters,  $k$ , and the AIC difference  $\Delta$ AIC from the best model.

**Appendix S4.** Maximum Likelihood IQ tree.  $n = 140$  individuals, 525 ddRAD loci = 56,515 sites (including 8861 variant sites) present in all populations with less than 60% missing data. Best substitution model TVMe+I + R3. 1000 UltrafastBootstraps.

**Appendix S5.** Detection of introgression with both hybrid species included in the dataset.

**Appendix S6.** AIC comparison for each model and species simulated under fastsimcoal2. The best scenarios ( $\Delta$ AIC < 5) are indicated in bold. Letters in the ‘Species’ column indicate which species is implied (A = *C. arcania*, G = *C. gardetta*, L = *C. leander*, H = *C. hero*). SI: Strict Isolation; IM: Isolation with Migration; SC: Secondary Contact; AM: Ancient Migration.

**Appendix S7.** Schematic representation of the twelve 4-species scenarios simulated with fastsimcoal2. Letters under each branch indicate the species (A = *C. arcania*, G = *C. gardetta*, L = *C. leander*, H = *C. hero*).

**How to cite this article:** Camizuli, T., Greenwood, M., Rioux, D., Zakharova, E. & Després, L. (2025) Evolutionary history and interspecific gene flow in six *Coenonympha* butterfly species in Europe. *Insect Conservation and Diversity*, 1–12. Available from: <https://doi.org/10.1111/icad.70005>