Hybridization promotes speciation in *Coenonympha* butterflies

Thibaut Capblancq, Laurence Despres, Delphine Rioux, Jesus Mavarez

► To cite this version:

Thibaut Capblancq, Laurence Despres, Delphine Rioux, Jesus Mavarez. Hybridization promotes speciation in *Coenonympha* butterflies. *Molecular Ecology*, 2015, 24 (24), pp.6209-6222. 10.1111/mec.13479 . hal-02464098

HAL Id: hal-02464098

<https://hal.science/hal-02464098>

Submitted on 15 Jan 2021

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

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

Hybridization promotes speciation in *Coenonympha* butterflies

Thibaut Capblancq^{1,2*}, Laurence Després^{1,2}, Delphine Rioux^{1,2} and Jesús Mavárez^{1,2}

¹ *Université Grenoble Alpes, LECA, F-38000 Grenoble, France*

² *CNRS, LECA, F-38000 Grenoble, France*

Date: July 29th, 2015

Running title: Hybridization and speciation in alpine butterflies

* Corresponding author: Thibaut Capblancq, 2233 Rue de la Piscine, 38041 Grenoble Cedex, France.

Tel: +33 (0)4 76 51 36 71, E-mail: thibaut.capblancq@gmail.com

ABSTRACT:

Hybridization has become a central element in theories of animal evolution during the last decade. New methods in population genomics and statistical model testing now allow the disentangling of the complexity that hybridization brings into key evolutionary processes such as local adaptation, colonization of new environments, species diversification and extinction. We evaluated the consequences of hybridization in a complex of three alpine butterflies in the genus *Coenonympha*, by combining morphological, genetic and ecological analyses. A series of Approximate Bayesian Computation procedures based on a large SNP dataset strongly suggest that the Darwin's Heath (*C. darwiniana*) originated through hybridization between the Pearly Heath (*C. arcania*) and the Alpine Heath (*C. gardetta*) with different parental contributions. The consequence of the hybridization is clear for Darwin's Heath morphology which is completely intermediate between the parental species whereas its climatic niche seems more similar to the Alpine Heath. Our results also reveal a substantial genetic and morphologic differentiation between the two geographically disjoint Darwin's Heath lineages leading us to propose the splitting of this taxon into two different species.

Keywords:

Hybrid species – Lepidoptera - ddRAD sequencing - Geometric morphometrics - Species diversification – Approximate Bayesian Computation

INTRODUCTION:

Hybridization between well-established species is known to be frequent in nature (Rieseberg 1997, Mallet 2007, Mavarez and Linares 2008, Schwenk *et al.* 2008, Descimon and Mallet 2010, Abbott *et al.* 2013), with around 10% of animal species and 25% of plants involved in interspecific crosses according to Mallet 2005. However, and particularly among zoologists, this process has for long been considered to be either a genetic dead end leading mainly to sterile hybrids or to have no major effect in the divergence of species, even when fertile hybrids are produced (Mayr 1963). Indeed, it is generally accepted that gene flow through hybridization between diverging lineages can delay the achievement of reproductive isolation by homogenizing the gene pools, working against the speciation process (Elgvin *et al.* 2011, Yakimowski and Rieseberg 2014). These views have been challenged during the last decades and interspecific hybridization is now accepted as a major process with a profound impact on several important evolutionary phenomena such as adaptation (Gompert *et al.* 2008, Kim *et al.* 2008, Song *et al.* 2011, The Heliconius Genome Consortium 2012, Pardo-Diaz *et al.* 2012), extinction (Rhymer and Simberloff 1996, Vonlanthen *et al.* 2012, Hohenlohe *et al.* 2010) and diversification/speciation (Mallet 2007, Mavarez and Linares 2008, Nolte and Tautz 2010, Nice *et al.* 2013, Yakimowski and Riesberg 2014). In particular, there is a recent interest in studying the contribution of hybridization to speciation through the formation of new hybrid taxa. This has led to the identification of hybrid species in diverse animal groups including insects (Mavarez *et al.* 2006, Salazar *et al.* 2010, Kunte *et al.* 2011), birds (Brelsford *et al.* 2011, Hermansen *et al.* 2011), fishes (Nolte *et al.* 2005) and mammals (Amaral *et al.* 2014). These studies, mostly conducted during the

last decade, have contributed to our understanding of hybrid speciation, which is defined as the emergence of a different, stable and isolated population/species by hybridization between two different taxa (Rieseberg 1997, Gross and Rieseberg 2005, Seehausen 2004, Mallet 2007, Mavarez and Linares 2008, Nolte and Tautz 2010, Abbott *et al.* 2013). A distinction has been made between “hybrid speciation”, where the emergence and isolation of the newborn lineage are the direct consequences of hybridization and “secondary gene flow”, where a lineage already diverged from a species undergoes introgression from another one (Mavarez and Linares 2008, Schumer *et al.* 2014). The major difference is the role played by hybridization during the initial divergence of the third species. However, even if the evolutionary steps differ, both processes can result in the stabilization of a novel lineage whose genome is a composite of two parental contributions and that can therefore be qualified as a “hybrid species”.

Hybrid species appear to be quite rare in nature, but the expanding literature on this topic is beginning to provide valuable information regarding the patterns associated with this phenomenon (Seehausen 2004, Mallet 2007, Mavarez and Linares 2008, Abbott *et al.* 2013). Some studies have highlighted a primordial role of intermediate morphology in the emergence of the newborn taxon (Mavarez *et al.* 2006, Salazar *et al.* 2010), while others have focused more on the crucial ecological adaptations brought by the genetic hybrid combination (Nolte *et al.* 2005, Schwarz *et al.* 2005, Gompert *et al.* 2006, Kunte *et al.* 2011) or the intrinsic genomic mechanisms which have allowed the isolation of the hybrid lineage from parental species (Lukhtanov *et al.* 2015). Taken together, these studies suggest that even if population genetics provide definitely the best tools to identify hybrid species, multiple lines of evidence, including morphology,

ecology and genetics must be taken into account when investigating the hybrid origin of a taxon. Furthermore, most studies to date have proposed the hybrid speciation hypothesis for a lineage based on its patterns of genetic composition and/or phenotype intermediacy, without further statistical testing. Yet recent improvements in DNA sequencing (e.g. genome projects, large SNP datasets, etc.) and model testing (e.g. Approximate Bayesian Computation procedures) allow for more powerful testing of the specific evolutionary processes and the scenarios involved during a hybrid speciation event (Nice *et al.* 2013, Sun *et al.* 2014).

Here we examine the role of hybridization in the evolutionary history of three European butterflies in the genus *Coenonympha* (Nymphalidae: Satyrinae). Two are well differentiated species: a) the Pearly Heath (*C. arcania*, Linné 1761), a common lowland butterfly present from Spain to Russia and b) the Alpine Heath (*C. gardetta*, Prunner 1798) a species living above 1500m with a disjoint distribution including mainly the Alps, but also the French massif central, Bosnia and Herzegovina, and Albania (Tolman and Lewington 2009). These two species diverged 1.5 to 4 million years ago according to a phylogeny of the genus *Coenonympha* based on mitochondrial *COI* sequences (Kodandaramaiah and Wahlberg 2009). The third taxon, the Darwin's Heath (*C. darwiniana*) shows an intermediate morphology and completely replaces *C. gardetta* in two distinct and narrow areas: the Valais and Tessin regions in Switzerland and the South West of the French Alps. Limited contact zones between pairs of the three taxa have been reported (Porter *et al.* 1995) but on-going gene flow between them is unknown. The taxonomic status of *C. darwiniana* has been hotly debated and specialists have variously classified it as a subspecies of *C. arcania*, a subspecies of *C. gardetta* or as hybrid between them (Porter *et al.* 1994, Wiemers *et al.* 1998, Schmitt and Besold 2010,

Lafranchis 2000, Tolman and Lewington 2009). However, most of these claims were based on studies with limited geographic coverage, without proper analyses of the ecology and morphology of the species and with genetic analyses based exclusively on allozyme markers (Porter *et al.* 1994, Wiemers *et al.* 1998, Schmitt and Besold 2010). To resolve the evolutionary history of the complex we first use a large SNPs dataset to compare different scenarios through ABC procedure. We especially tested for the hybrid origin of *C. darwiniana* by confronting this hypothesis with alternative evolutionary models of cladogenesis. As mentioned above, a combination of evidence (*i.e.* phenotypes, behaviors, ecology, genetics, etc.) is helpful to support the hybrid origin of a taxon and understand the context of such speciation event. Yet this kind of analysis has never been made for *C. darwiniana*. We combined analysis based on population genetics, morphometry of wing shapes and comparison of climatic niches to identify the genetic, morphological and ecological patterns associated with the species and put them in regard to the scenario retained by the ABC analysis. This set of analyzes reveals new insights on the crucial role that hybridization has played in the diversification of this group of European butterflies.

MATERIAL AND METHODS:

Butterfly sampling

We sampled populations across the entire geographic range of *C. darwiniana* in the south French Alps and the Tessin region in Switzerland/Italy, as much as possible from the main range of *C. gardetta* (Alps), altogether with nearby populations of *C. arcania* within these areas. Using entomological nets, twelve populations of *C. darwiniana* (seven in South French Alps and five in Tessin) were sampled; we also

sampled eleven populations of *C. arcania* and nineteen populations of *C. gardetta* (Fig 1). Sample sizes per population depend on the analysis (genetics or morphology) and are listed in Table S1 (supporting information).

Genetic data acquisition

DNA was extracted from the complete thorax of each individual using the DNeasy Blood and Tissue Kit (QIAgen, Germany) according to the manufacturer's instructions and stored at -20°C. A double-digested RAD (Restriction site Associated DNA) experiment was conducted on 174 individuals using a modified version of the protocol described in Peterson *et al.* (2012). In short, 300ng of DNA template from each individual were double-digested with 10 units each of *Sbf*I-HF and *Msp*I (New England Biolabs Inc.) at 37°C during one hour in a final volume of 34µl using the CutSmart buffer provided with the enzymes. Digestion was further continued together with ligation of P1 and P2 adapters (see Peterson *et al.* 2012) by adding 10 units of T4 DNA ligase (New England Biolabs Inc.), adapters P1 and P2 in 10-fold excess (compared to the estimated number of restriction fragments) and 1µl of 10mM ribo-ATP (New England Biolabs Inc.) in each sample. This simultaneous digestion-ligation reaction was performed on a thermocycler using a succession of 60 cycles of 2 min at 37°C for digestion and 2 min at 16°C for ligation, after which the enzymes were heat inactivated at 65°C for 10 minutes. An equal volume mixture of all the digested-ligated fragments were purified with Agencourt AMPure XP beads (Beckman Coulter, France). Fragments were size-selected between 250 and 500bp by gel extraction (agarose gel: 1.6%) and purified with QIAquick Gel Extraction Kit (Qiagen, Germany). The ddRAD library obtained was amplified using the following PCR mix: a volume of 20µl with 1µl of DNA template, 10

mM of dNTPs, 10 μ M of each PCR probe (Paterson *et al.* 2012) and 2U/ μ l of *Taq* Phusion-HF (New England Biolabs Inc.); and the following PCR cycles: an initial denaturation at 98°C for 10 minutes; 15 cycles of 98°C for 10 sec, 66°C for 30 sec and 72°C for 1 min; followed by a final extension period at 72°C for 10 min. The amplified ddRAD library was purified with QIAgen MinElute PCR Purification Kit (Qiagen, Germany) and sequenced on an Illumina Hi-Seq 2000 (4/10 lane, paired-end 2 x 100bp, Fasteris SA, Switzerland).

The DNA reads (120 million reads) from the library were used to call SNP genotypes with the *STACKS* pipeline (Catchen *et al.* 2013) to create the SNPs dataset used in the ABC procedure (see section “*Test of evolutionary history scenarios with Approximate Bayesian Computation (ABC) method*”) and multivariate analysis (see section “*Multivariate analysis*”).

Test of evolutionary history scenarios with Approximate Bayesian Computation (ABC) method

We used the *STACKS* pipeline to treat the ddRAD sequencing data as follows: the *process_radtags* function was first run to demultiplex the data and filter the reads on their quality with a sliding window of 15% of the read length and a Phred score limit of 10. Reads with low quality or uncalled bases were removed. A *de novo* assembly was then performed on each individual read pool using the *ustacks* function, the minimum coverage for creating a stack of identical reads (-m) was fixed to 5 and a maximum of 4 different nucleotides was imposed to merge two different stacks in a polymorphic locus (-M). Highly-repetitive stacks and over merged tags were dropped using both the “Removal algorithm” (-r) and the “Deleveraging algorithm” (-d). A data set of the loci

from all the individuals' loci data sets was then built using the *cstacks* function in order to find all the alleles in the libraries and to produce a catalog of consensus alleles for each locus. A maximum of 5 mismatches was allowed for considering two individual tags as the same locus and to merge them in the catalogue (-n). The sets of loci constructed within each individual read pool were then searched against the newly produced catalog (*sstacks* function) and a SNP dataset was produced with the genotype of each individual for every polymorphic position (*populations* function). Only tags present in more than 60% of the sampling were used to find the SNPs in order to avoid an excess of missing data. Finally, SNPs with a frequency lower than 5% were removed from the data set and only one polymorphic site was kept for each RAD-tag. This genetic matrix was then used to test different models of evolution.

Several scenarios have been proposed for the evolutionary history of these three *Coenonympha* species and especially for the origin of *C. darwiniana*. To compare them and find the most plausible one we used the SNP dataset with the Approximate Bayesian Computation (ABC) approach implemented in the software DIYABC (Cornuet *et al.* 2014). In short, a large number of simulated data sets were produced under each tested scenario by sampling parameter values into predefined prior distributions and an estimation of each scenario posterior probability was calculated as a function of the similarity between several summary statistics of simulated data sets and observed data using a logistic regression procedure. These posterior probabilities enable a ranking of scenarios and allow for the identification of the most probable evolutionary history. We only considered five evolutionary scenarios (Fig 2), based on previous published studies dealing with the evolution of the complex (Porter *et al.* 1995, Wiemers *et al.* 1998 and

Schmitt and Besold 2010) and our own hypothesis based on the morphometric and phylogenetic analyses (see Results).

Scenario design and simulations: A uniform distribution with a large interval was chosen for each prior increasing the need in computation time but overcoming the lack of knowledge on population sizes, divergence times and admixture rate (see Tab. S2 in Supporting information). Only one approximate *C. arcania/C. gardetta* divergence time of 1.5-4 million of years was available from Kodandaramaiah and Wahlberg (2009). For each scenario a total of one million data sets were simulated and the posterior probability was computed by performing a logistic regression on the 1% of simulated data closest to the observed data set (Cornuet *et al.* 2008, 2010). Summary statistics used for observed/simulated datasets comparisons are the mean gene diversity and Fst across all loci and Nei's distances among populations. Gene diversity was chosen because allelic richness is directly influenced by hybridization; the genetic differentiation (Fst) and distance (Nei's) among taxa were used because these indexes reflect the degree of differentiation between populations and are highly relevant to compare evolutionary histories.

Assessing the "goodness-of-fit" of the model: Once the most probable scenario was identified, different posterior analyses were done to evaluate the trustfulness of the procedure. First, to be sure of the choice of the scenario, 500 new datasets were simulated with each historical model creating pseudo-observed datasets for which posterior probabilities were re-estimated with the same procedure described above. Type I and type II errors were evaluated by measuring, respectively, the fraction of datasets simulated under the best scenario that were assigned to other scenarios and

the fraction of datasets simulated under other scenarios that were assigned to the best scenario. Second, we wanted to know if our selected model is able to produce datasets similar to the observed one. This confidence in the model is evaluated by estimating the similarity between simulated (1000 simulations) and real datasets using summary statistics different than the ones used for model choice (an important precaution, see Cornuet *et al.* 2010). For each summary statistic a *P*-value is estimated by ranking the observed value among the values obtained with simulated datasets and a principal component analysis was also performed to check visually the position of the observed data in relation to the data sets generated from the posterior predictive distribution.

Posterior distribution of parameters and their precision: It was possible to evaluate the posterior distributions of parameters for our selected scenario using a local linear regression on the 1% of simulations closest to the observed data set (Cornuet *et al.* 2010). These distributions gave an idea of the most probable value (median) and the “approximation” (width of the distribution) for each historical model. A precision of these estimations was evaluated by computing the median of the absolute error divided by the true parameter value of the 500 pseudo-observed data sets simulated under the selected scenario using the median of the posterior distribution as point estimate (Cornuet *et al.* 2010)

Multivariate analysis

We investigate individuals grouping and differentiation without any assumptions through a Principal Component Analysis (PCA) carried out on the same SNPs dataset. We used R 3.1.0 (R Development Core Team) and the library adeGenet (Jombart 2015)

to achieve the PCA. Missing values were replaced by the mean frequency of the concerned allele in all the sampling.

Geometric morphometry analysis

We performed a geometric morphometric analysis including wing shape, patterns and venation, with a particular interest on traits that have been previously considered as indicative of intermediacy in *C. darwiniana* such as the form and position of eyespots, the shape of the hindwing white band and the global size of the wings. We identified 22 and 18 homologous landmarks in the fore- and hindwings, respectively, corresponding either to: a) “pure” wing shape when using vein intersections and termini or: b) wing patterns when using eyespots and the white band outline (see Fig. S3 in Supporting information).

The wings were scanned with a resolution of 1200 dpi and the geometric morphometric analysis was performed using the tps (thin-plate spline) series of programs (available at <http://life.bio.sunysb.edu/morph/>) for file creation (TPSUTIL, ver 1.58, Rohlf 2006b) and landmarks positioning (TPSDIG 2, ver 2.17, Rohlf 2006a). A procustes superimposition was then applied to the landmarks coordinates in order to remove non-shape variation in location, scale and orientation. After a dimensionality reduction to correct for the effect of using a large number of variables relative to the number of specimens was applied (see Mérot *et al.* 2013), we tested shape differences among species using a one-way MANOVA on the remaining principal components (PCs) with shape as dependent variable and species as factor. A canonical variate analysis (CVA) allowed for the investigation of species discrimination (Mérot *et al.* 2013), and a dendrogram was produced with morphometric distances (Mahalanobis distances)

among groups to allow for a clearer visualization of population morphological position in the three species complex. The generalized procustes analysis and the different statistics were performed with R and the Rmorph (Baylac, 2012), ade4 (Chessel, Dufour and Thioulouse, 2004) and APE (Paradis, Claude and Strimmer, 2004) libraries. In parallel to the Procrustes superimposition and shape analysis, a wing size index was obtained using the log-transformed centroid size directly from the wing landmarks. Difference in size among species was tested through an analysis of variance (ANOVA) with size as a dependent variable and species as factor.

Climatic niche comparisons

Occurrence data of the three species in the Alps were collected from numerous institutions datasets: FLAVIA (French association of entomology, <http://www.flavia.ade.free.fr>), MNHN (French national museum of natural history, <http://www.mnhn.fr>), CSCF (Centre Suisse de Cartographie de la Faune, <http://www.cscf.ch>), Mercantour National Park (<http://www.mercantour.eu>), Asters CEN (Conservatoire d'Espaces Naturels de haute Savoie, <http://www.asters.asso.fr>), LPO (Ligue pour la protection des oiseaux, <http://www.lpo.fr>), GBIF (Global Biodiversity Information Facility, <http://www.gbif.org>) and personal data from Michel Savourey and Roger Vila. After deleting all points with too low precision of coordinates or with obvious mistakes, around 6900 points remained for *C. arcania*, 4000 points for *C. gardetta* and 1000 for *C. darwiniana* distributed all along the Alps and adjacent regions. These occurrence data were associated to the 19 "Bioclimatic" variables available in the WORLDCLIM database (Hijmans *et al.* 2005, <http://worldclim.org>). These variables represent a range of metrics (mean value, variance, and extremes) on global

temperature and precipitations. For more precision we used the 30 seconds resolution grids ($\approx 1\text{km}^2$) and we kept only one occurrence per species in each pixel in order to avoid overweighed values. Importation and manipulation of the data were done in R (R Development Core Team) using *dismo* and *raster* libraries (Hijmans *et al.* 2015).

An ordination method “outlying mean index” (OMI, Dolédec *et al.* 2000) was used to compare each species environmental niches obtained from occurrence data and climatic variable correspondence. This approach gives the position and the breadth of each species niche along main axes calculated from a set of environmental variables. The niche position corresponds to the mean location of the species in the environmental space and the breath represents the variability of the used environment (SD of the species values in the OMI axes, Thuiller *et al.* 2004). The method is appropriate to evaluate ecological niche proximity among taxa and gives an idea of the environmental capacities of each species. The analysis was performed in R using the *ade4* library.

RESULTS:

Genetic evidence of hybridization

The double-digested RAD libraries produced a mean of 3200 fragments (RAD-tags) with a mean coverage of 35 reads/fragment for the 174 individuals analyzed. The polymorphism in these RAD-tags allowed the scoring of 851 independent SNPs. These markers were first used for the evolutionary model testing with the Approximate Bayesian Computation procedure and the results obtained are summarized in the Fig 3. The posterior probabilities (Fig 3A) show that the scenario involving hybridization between *C. arcania* and *C. gardetta* at the origin of an ancestral population of *C.*

darwiniana (scenario 4) is far superior to any other scenario. All scenarios involving only cladogenesis without hybridization appeared highly improbable given the observed data set and were thus completely rejected by the analysis. Scenario 4 supposes a unique ancestral population for *C. darwiniana* and thus a common origin for the taxon, which would have split later into the two geographical clusters present today (Fig 3B). It is worth mentioning that the genetic differentiation and distinct geographical range of these two *C. darwiniana* clusters prompted us to test a scenario with two independent hybridization events for the origins of these two entities (scenario 5). However, it also showed a posterior probability close to zero and was therefore rejected.

It is relevant to observe that scenario 4 is the one with the best fit (40% of the summary statistics calculated on datasets simulated with scenario 4 are very close to the summary statistics calculated with the real genetic dataset) in comparison with the other four scenarios tested (only 19 to 33%). Our model produces simulations slightly distant from the observed genetic dataset in the summary statistics space, but given the simplicity of designed scenarios, it is not surprising that they do not perfectly match the real history of our populations and do not produce simulations in complete concordance with observed data. Moreover, type I (11.4%) and type II (0.05%) errors were very low in comparison with similar studies (Barres *et al.* 2012, Cornuet *et al.* 2010). The scenario 4 was thus greatly distinguishable from the others. Taken together these scores in post ABC procedure verifications strongly support the evolutionary scenario 4, *i.e.* a hybridization event, at the origin of *C. darwiniana*.

All the parameter estimates under the selected scenario show quite realistic values (Fig 3), with a divergence time between *C. arcania* and *C. gardetta* of 1.7 ± 0.05

Myr (t_3 in Figure 3C). The time for the hybridization event that gave origin to *C. darwiniana* (t_2 : 12000 $\pm$ 100 years) and the time for the splitting between the two geographical populations currently observed in this species (t_1 : 10000 $\pm$ 150 years), appear nonetheless quite recent. However, these two times correspond to a suitable period for numerous *C. arcania*/*C. gardetta* contacts and species range upheaval (see Discussion). Our analysis also indicates that the initial parental contributions in the genetic composition of the ancestral *C. darwiniana* population (*i.e.* parameters ra and $1-ra$) were not 50/50. Indeed, there is evidence for larger contribution of a *C. arcania* genetic background (75%), suggesting multiple events of hybridization and backcrossing between the hybrid and the parental species.

PCA analysis identifies four well differentiated clusters, corresponding to the putative parental species *C. arcania* and *C. gardetta*, and the two geographic isolates of *C. darwiniana* (Fig 4). The *C. arcania* and *C. gardetta* groups exhibit the highest inter-species differentiation along the most informative principal component (PC1 19.3%), while the two geographical groups of *C. darwiniana* take intermediate scores along the axis. The second PC (7.3%) differentiates these two populations of *C. darwiniana* from the other two species and points out the genetic variability intrinsic to these lineages. Finally, we observe that the *C. darwiniana* individuals are clustered in two distinct groups along the third PC (3.7%), corresponding to the populations from the Tessin area and the south France Alps, respectively. A substantial genetic structure within *C. darwiniana* is thus highlighted by this analysis and matches perfectly the discontinuous geographic distribution of this species.

Morphological proximities among the complex

A total of 510 specimens had both right fore- and right hindwings in good conditions and were included in the morphological analysis. A major result of the morphometric analysis is the good concordance they show with the genetic clustering analysis (Fig 5A). We found highly significant differences in size (ANOVA: p-value < 2e-16, df = 3, F-value = 172.2) and shape (MANOVA: p-value < 2e-16, df = 3, approx. F-value = 14.5) among the four previously established genetic clusters. For these two traits the strongest differentiation is visible between *C. arcania* and *C. gardetta* individuals with almost no overlap in wing centroid size values (Fig 5B) and in scores of the discriminant analysis first axis (Fig 5A). This first axis of the CVA combines the majority of the discriminant variability (80%) and separates clearly the four groups, with both *C. darwiniana* populations taking intermediate values between the two putative parental species (Fig 5A-C). Other analyses of landmarks corresponding only to either venation, the shape of the hindwing white band or only the positions of hindwing eyespots confirm this intermediacy (not shown). Our results robustly support the intermediate position of *C. darwiniana* in regards to the putative parental species, not only for easily visible characters in the phenotypes such as the eyespots' color or their alignment, as previously highlighted by some experts, but also for most of the wing structures.

The analysis also showed a significant distinction between the two *C. darwiniana* populations (MANOVA, p-value = 1.4e-07, approx. F-value = 4.6), in agreement with the results from the genetic differentiation analyses. Furthermore, we observe that although in general *C. darwiniana* populations appear intermediate between parental species, *C. darwiniana* from Switzerland shows a closer phenotypic proximity with *C. gardetta* as

shown by the large overlap between these two groups in the discriminant analysis result (Fig 5A).

Finally, the relative homogeneity of each group, in terms of wing shape and size, points out their morphological integrity. The second axis of the CVA corresponds to *C. darwiniana* specific variability, which emphasizes the differentiation within this taxon (two distinct genetic lineages) even if it concerns only a little part of the discriminant information (LD2, 14%).

Climatic niches analysis

The climatic niches of the four populations can be visually inspected in Fig 6. The first two axes of the OMI analysis explained almost all the variability of climatic niches and represented respectively 96 (OMI1) and 3% (OMI2) of the total inertia. The first one, negatively correlated to the rainfall variables and positively correlated with the temperature variables, mostly corresponds to altitudinal effects with a part of local specificities. It thus logically separates *C. arcania* and *C. gardetta* with great significance for each climatic variable. On this axis both French and Swiss *C. darwiniana* appear to have a climatic niche much closer to *C. gardetta*, with only French *C. darwiniana* showing a slight overlap with *C. arcania*. This is concordant with the fact that *C. darwiniana* completely replaces *C. gardetta* at slightly lower elevations in southern latitudes. On the other hand, the fact that Swiss *C. darwiniana* shows a higher degree of overlap with *C. gardetta* for climatic conditions supports, again, a closer proximity between these two taxa. Although the second axis (OMI2) only explains 3% of the variability, it distinguishes the humid continental climate influence in the central and east Alps with cool and rainy summers (high mean T°C values during the wettest quarter) and the

oceanic influence in the western flank of the Alps with more precipitation during the winter and spring (high mean T°C values during the driest quarter). This axis mainly differentiates the two *C. darwiniana* taxa, with the Swiss population in continental climatic conditions and the French population in mountain ranges with lower summer rainfall.

DISCUSSION:

Hybrid species hypothesis for Coenonympha darwiniana

By comparing different evolutionary scenarios based on numerous genetic markers we validated the hybrid origin hypothesis for the *C. darwiniana* lineages and were able to partially reconstruct the history of the complex this taxon forms with its parental species *C. arcania* and *C. gardetta*. The ddRAD sequencing technique allowed us to genotype a large number of co-dominant markers (851 SNPs) which are likely widely distributed over the species genomes (one of the advantages of the RADseq and similar techniques, see Baird *et al.* 2008). This property of the markers was ideal for quantifying the posterior probability of different evolutionary histories through Approximate Bayesian Computation methods. This procedure clearly favored the hybrid species scenario giving it by far the highest probability among the five evolutionary models tested. ABC methods have already been used to compare different types of evolutionary scenarios (Estoup *et al.* 2004, Csillery *et al.* 2010, Barres *et al.* 2012), including the study of hybridization events in another butterfly species (Nice *et al.* 2013). But keeping in mind that no abstract models can fully account for the inherent complexity of species evolutionary history (Csillery *et al.* 2010) and that ABC approaches are sensitive to complexity differences in the compared models, we decided to use this approach only to

choose between three pure cladogenesis scenarios and two simple scenarios combining cladogenesis and hybridization (see Fig 2). These scenarios correspond to the most credible evolutionary steps that could have led to the currently observed genetic structure among the three species. The results, completely rejecting the scenarios involving only cladogenesis events, provide strong evidence in favor of the hypothesis that hybridization was involved in the evolutionary events that gave rise to *C. darwiniana* and its resulting patterns of genetic, morphological and somewhat ecological intermediacy. Moreover, it was possible to distinguish between two relatively close alternatives: a single hybridization event implying a unique origin and a double hybridization event implying two independent origins for the two *C. darwiniana* lineages. The unique hybrid origin was strongly preferred during the analysis with a differentiation of the two populations coming only later. We acknowledge, however, that the simplicity of the tested scenarios do not permit for a more precise reconstruction of all the steps in the evolutionary history of the complex. Such a detailed reconstruction was not the objective of this work.

The results provided by the multi-proxy approach are in agreement with the retained scenario and offers substantial information about the consequences of hybridization on the phenotypes and ecology of the *C. darwiniana* lineages. The potential implications of hybridization in species evolution remains little known (Mallet 2005, Abbott *et al.* 2013), yet the genetic, morphological and ecological patterns associated with a hybrid species can provide new insights to better understand the possible outcomes of the phenomenon:

Genetics: PCA analysis provides an idea of genetic clustering without making any assumptions and supports the hybrid origin scenario. The two *C. darwiniana* lineages have intermediate scores between *C. arcania* and *C. gardetta* along the first PC which reflects the highest genetic structure within the complex. Although not a formal proof of hybrid origin, this result greatly concords with that hypothesis. The successive differentiation shown by the first three PCs completely correspond to the successive events of the hybrid species scenario (parental differentiation, hybridization event and hybrid lineages differentiation, see Figure 3). As stated above, the first PC shows the ancestral divergence of the lowland and alpine species with intermediacy of the two lineages of *C. darwiniana*. The second axis corresponds to the distinction of the hybrid lineages on one side and the parental species on the other side, showing the intrinsic genetic variability of these taxa and their history since the speciation event. Finally, PC3 discriminates the two *C. darwiniana* populations, consistent with the geographic location of the individuals. This result increases our confidence in our conclusions and adds support to the view that ddRAD sequencing can be a powerful technique for genetic clustering at a fine scale (Wagner *et al.* 2012, Rubin *et al.* 2012, Cruaud *et al.* 2014).

If we accept a common and unique origin for the two *C. darwiniana* populations, which is supported by the ABC method, it is surprising to have two distinct groups in the PCA analysis (Fig 3). Although both taxa show intermediate patterns when compared with the parental species, French *C. darwiniana* is closer to *C. arcania* whereas Swiss *C. darwiniana* is closer to *C. gardetta* (similar patterns in morphology and climatic conditions). This observation first led us to suppose two independent hybrid lineages as

shown in North American *Lycaeides* butterflies (Gompert *et al.* 2009, Nice *et al.* 2013). But the ABC modeling clearly rejected this possibility and converged toward a unique and common origin for the two *C. darwiniana* taxa. Given the recent origin of the taxon (< 20000 years), the divergence observed between them seems large to be due only to independent evolution after geographical splitting. A possible explanation would be a differential backcrossing rate with the parental species for each of the two *C. darwiniana* lineages. Under this hypothesis, independent parental genetic contributions posterior to the lineage splitting observed within *C. darwiniana* would explain the substantial separation. The genetic stability evident now in the populations would be the result of their repeated interactions with one or both parental species, possibly driven by different environmental pressures. However, this hypothetical scenario remains uncertain and needs to be tested. It could provide an interesting case of successive hybridization events leading to differential adaptations, and illustrate the tenuous equilibrium between backcrossing due to incomplete reproductive isolation of the newborn lineage and success of hybrid population establishment.

Morphology: Morphology can play an important role in species isolation for butterflies and other animals because of visual preferences associated with mating behaviors. This is one of the reasons for which it has been investigated for most animal hybrid species. In some cases the hybrid lineage shows extreme traits compared to the parental species (e.g. a sculpin lineage in Nolte *et al.* 2005 or the bat *Artibeus schwartzi* in Larsen *et al.* 2010), with these extreme phenotypes being potentially advantageous in species fitness. But most hybrid species present an intermediate morphology (e.g. the hybrid butterfly *Heliconius heurippa* in Mavarez *et al.* 2006 and Salazar *et al.* 2010, the birds *Passer*

italiae in Hermansen *et al.* 2011, *Dendroica auduboni* in Brelsford *et al.* 2011 or the butterfly *Papilio appalachiensis* in Kunte *et al.* 2011). More than just being a consequence of hybridization, intermediate traits could play a major role in species emergence and rise, as was the case for *Heliconius heurippa* species (Mavarez *et al.* 2006, Salazar *et al.* 2010). In our case we found no transgressive traits but a strong morphological intermediacy for most of the wing elements, including patterns and venation shape, positioning and size. This morphology of *C. darwiniana*, easily visible when looking at specimens of each species, has been the principal argument behind the hypothesis of its hybrid origin (Porter *et al.* 1995). The possibility that intermediate values in wing traits are independent from hybridization for this taxon seems unlikely when taking all the evidence into account. Although it could be only a neutral consequence of hybridization, the homogeneity of taxa in terms of wing patterns lead us to suspect that the intermediacy could be involved in species isolation. Intermediate phenotype preferences by hybrids have been recorded for fish species (Selz *et al.* 2014) and could have been involved in *C. darwiniana* rise. Furthermore, the two *C. darwiniana* lineages differ from each other but within each population the morphology is quite homogenous. Intra-*C. darwiniana* morphological variability is not as large as the morphological variability we encountered in the parental species taken together. From all the phenotypic combinations potentially brought by the mixture of the parental entities, the newborn lineage probably only rose and established from a restricted pool of them.

Ecology: Ecology is known to play a major role in hybrid species emergence (Mallet 2007, Mavarez and Linares 2008, Abbott *et al.* 2013). The mix of different genetic

backgrounds leads to novel combinations that can favor the rise of a hybrid lineage by multiple means (Seehausen 2004, Mallet 2007, Abbott *et al.* 2010, Abbott *et al.* 2013) including intermediate phenotypes (Mavarez *et al.* 2006, Kunte *et al.* 2011, Hermansen *et al.* 2011) and/or transgressive phenotypes (Nolte *et al.* 2005, Gompert *et al.* 2006). For example, one of the main mechanisms by which new hybrid lineages achieve reproductive isolation *vis-à-vis* their parental species seems to be the colonization of habitats not used by them (Rieseberg 1997, Nolte *et al.* 2005, Gompert *et al.* 2008, Nice *et al.* 2013). We investigated the positioning of the hybrid species in terms of climatic conditions to determine whether an extreme climatic niche was a key parameter in the hybrid species' emergence. However, we found that the climatic niches of *C. gardetta* and *C. darwiniana* strongly overlap, especially for the Swiss *C. darwiniana* lineage, although the two species only share narrow sympatric zones in the wild. This indicates that large-scale climatic conditions are unlikely to be responsible for the differences in the geographic distributions of these species. Temperature and precipitation regimes were probably not the mechanism behind species isolation. Nonetheless, the climatic niche does appear intermediate for the French *C. darwiniana* lineage, supporting the view of some degree of adaptation to intermediate climatic conditions. That is not senseless considering the southern exposure and modest altitudes of the mountain ranges where they are found in France. Further analyses of ecological differentiation, such as fine-scale habitat use and relationships with host-plants, remain necessary to paint a clear picture of the role of ecology in the differentiation, isolation and coexistence of both *C. darwiniana* lineages *vis à vis* their parental species.

Our study provides strong evidence of hybridization between two distinct *Coenonympha* species in the composition of *C. darwiniana* and a substantial reproductive isolation as suggested by the genetic, morphologic and climatic clustering of the taxa. However, an important additional aspect needs to be evaluated: the evidence that reproductive isolation between the hybrid and parental lineages is a direct consequence of hybridization (Mavarez and Linares 2008, Nolte and Tautz 2010, Schumer *et al.* 2014). It is difficult with our results to determine whether hybridization has been the key mechanism of the speciation process (as shown for example for a sculpin hybrid species in Nolte *et al.* 2005, or for a hybrid butterfly species in Mavarez *et al.* 2006) or whether a lineage already diverged from one of the parental species has undergone a strong introgression from the other. However, the current stability, homogeneity and specificities of *C. darwiniana* in its two geographic populations together with the complete exclusion of parental species in these regions suggests an important role of hybrid patterns in the adaptation and persistence of this taxon.

Structure of the complex and taxonomic considerations

The genetic and morphological analyses differentiated four entities within this species complex with complete concordance and in agreement with the “four units complex” hypothesis first proposed for this group by Schmitt and Besold (2010). For each of the analyses used the clustering pattern was similar, even if the climatic analysis failed to clearly discriminate the three alpine taxa (*C. gardetta* and the two *C. darwiniana* populations). The main differentiation was observed between *C. arcania* and *C. gardetta*, with no overlap between these two species regardless of which marker was used. Contrastingly, the *C. darwiniana* individuals clustered in two groups corresponding to

the two distinct geographical ranges of the taxon, one in the Tessin area (mostly in Switzerland) and the other in the southwest of the Alps (mostly in France). Furthermore, each of the four “units” shows intrinsic homogeneity in their genetic composition, wing patterns and ecological (climatic) conditions. Regarding taxonomy, *C. arcania* and *C. gardetta* have long been considered “*bona species*” (Kodandaramaiah and Wahlberg 2009, Schmitt and Besold 2010, Descimon and Mallet 2009, Lafranchis 2000) and our combination of approaches confirms this taxonomic arrangement. The status of *C. darwiniana* has been more debated (see Porter *et al.* 1995, Wiemers *et al.* 1998, Schmitt and Besold 2010, Descimon and Mallet 2009). The genetic positioning of this taxon within the complex was previously examined based on protein electrophoresis and conclusions varied from a sub-species of *C. gardetta* (Porter *et al.* 1995, Wiemers *et al.* 1998) to a sub-species of *C. arcania* (Schmitt and Besold 2010). The small number of markers used prevented these studies from detecting the hybrid composition of *darwiniana*’s genetic background, and their divergent conclusions are consistent with the fact that hybridization can lead to incongruence in the coalescence patterns of genetic markers (Leaché and McGuire 2005, Meng and Kubatko 2008, Roos *et al.* 2011). Despite these considerations the taxon is now mostly considered as a different species (Lafranchis 2000, Tolman and Lewington 2008). The observed differentiation of *C. darwiniana* in regards to its two parental species effectively agrees with the “cluster” definition of a species (Coyne and Orr 2004, Mallet 2007). Even so, one question remains unanswered: how should we treat the two geographic populations of *C. darwiniana*? Although they are genetically, morphologically and climatically close, and appear to have a relatively recent common origin, we are definitely able to separate them in two distinct groups. For this reason we propose to go further than Schmitt and Besold

(2010) by giving species status to each of the two currently recognized sub-species of *C. darwiniana*. Their disjoint geographic ranges together with their differences in genetic composition, wing structure and patterns and, to a lesser extent, their climatic niches, seem to us as enough evidence to justify their species status. We propose to raise the names previously used for the subspecies (Schmitt and Besold 2010) to a species level: *Coenonympha darwiniana* for the Tessino population and *Coenonympha macromma* for the southwest Alps population.

Further studies are needed now to investigate the mechanisms (ecological, behavioral, or others) that maintain reproductive isolation among these four populations into the wild. Such information could greatly help to better understand the factors which have favored the hybrid lineages rise and provide evidence concerning the potential adaptive role of hybridization.

ACKNOWLEDGEMENTS:

We are very grateful to Flavia entomologist association, Mercantour National Parc, French National Natural History Museum and Asters conservatory for their help in data collection. We would also like to thank Yann Baillet, Pascal Dupont, Marie-France Leccia, Gregory Guichert and Eric Drouet for helping us collecting samples and giving us helpful advises on the project. Finally we thank Luis Daniel Llambi, Eric Bazin and Frédéric Boyer for comments and revision of the manuscript. This project was funded by the “Conseil general de l’Isère” and a “CNRS INEE – APEGE” grant.

REFERENCES:

- Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. *Taxon*, **59**, 1375–1386.
- Amaral AR, Lovewell G, Coelho MM, Amato G, Rosenbaum HC (2014) Hybrid Speciation in a Marine Mammal: The Clymene Dolphin (*Stenella clymene*). *PLoS ONE*, **9**.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.
- Barão KR, Gonçalves GL, Mielke OHH, Kronforst MR, Moreira GRP (2014) Species boundaries in *Philaethria* butterflies: an integrative taxonomic analysis based on genitalia ultrastructure, wing geometric morphometrics, DNA sequences, and amplified fragment length polymorphisms. *Zoological Journal of the Linnean Society*, **170**, 690–709.
- Barrès B, Carlier J, Seguin M *et al.* (2012) Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity*, **109**, 269–279.
- Brelsford A, Milá B, Irwin DE (2011) Hybrid origin of Audubon’s warbler. *Molecular Ecology*, **20**, 2380–2389.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
- Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package-I-One-table methods. *R news*, **4**, 5–10.

618 Consortium THG (2012) Butterfly genome reveals promiscuous exchange of mimicry
619 adaptations among species. *Nature*, **487**, 94–98.

620 Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make
621 approximate Bayesian computation inferences about population history using single
622 nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**,
623 1187–1189.

624 Cornuet J-M, Ravignie V, Estoup A (2010) Inference on population history and model
625 checking using DNA sequence and microsatellite data with the software DIYABC
626 (v1.0). *BMC Bioinformatics*, **11**, 401.

627 Cornuet J-M, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY
628 ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*,
629 **24**, 2713–2719.

630 Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philosophical*
631 *Transactions of the Royal Society of London B: Biological Sciences*, **353**, 287–305.

632 Cruaud A, Gautier M, Galan M *et al.* (2014) Empirical assessment of RAD sequencing for
633 interspecific phylogeny. *Molecular biology and evolution*, msu063.

634 Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian
635 Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

636 Descimon H, Mallet J (2009) Bad species. *Ecology and Evolution of European Butterflies*
637 (*Shreeve, TG and Dennis, RHL, eds*), Oxford University Press.

638 Dolédec S, Chessel D, Gimaret-Carpentier C (2000) Niche separation in community analysis:
639 a new method. *Ecology*, **81**, 2914–2927.

640 Dray S, Dufour A-B, Thioulouse J *et al.* (2013) *ade4: Analysis of Ecological Data :*
641 *Exploratory and Euclidean methods in Environmental sciences*. Available at

642 <http://cran.r-project.org/web/packages/ade4/index.html>.

643 Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses.

644 *Bioinformatics*, **30**, 1844–1849.

645 Elgvin TO, Hermansen JS, Fijarczyk A *et al.* (2011) Hybrid speciation in sparrows II: a role

646 for sex chromosomes? *Molecular Ecology*, **20**, 3823–3837.

647 Gautier M, Gharbi K, Cezard T *et al.* (2013) The effect of RAD allele dropout on the

648 estimation of genetic variation within and between populations. *Molecular Ecology*,

649 **22**, 3165–3178.

650 Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC (2006) Homoploid Hybrid

651 Speciation in an Extreme Habitat. *Science*, **314**, 1923–1925.

652 Gompert Z, Forister ML, Fordyce JA, Nice CC (2008) Widespread mito-nuclear discordance

653 with evidence for introgressive hybridization and selective sweeps in *Lycaeides*.

654 *Molecular Ecology*, **17**, 5231–5244.

655 Gross BL, Rieseberg LH (2005) The Ecological Genetics of Homoploid Hybrid Speciation.

656 *Journal of Heredity*, **96**, 241–252.

657 Hermansen JS, Sæther SA, Elgvin TO *et al.* (2011) Hybrid speciation in sparrows I:

658 phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular*

659 *Ecology*, **20**, 3812–3822.

660 Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

661 Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary.

662 *Philosophical Transactions of the Royal Society of London B: Biological Sciences*,

663 **359**, 183–195.

664 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution

665 interpolated climate surfaces for global land areas. *International Journal of*

666 *Climatology*, **25**, 1965–1978.

667 Hijmans RJ, Phillips S, Elith JL and J (2015) *dismo: Species Distribution Modeling*.
668 Available at <http://cran.r-project.org/web/packages/dismo/index.html>.

669 Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation
670 RAD sequencing identifies thousands of SNPs for assessing hybridization between
671 rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

672 Hughes CE, Atchison GW (2015) The ubiquity of alpine plant radiations: from the Andes to
673 the Hengduan Mountains. *New Phytologist*.

674 Kim M, Cui M-L, Cubas P *et al.* (2008) Regulatory Genes Control a Key Morphological and
675 Ecological Trait Transferred Between Species. *Science*, **322**, 1116–1119.

676 Kodandaramaiah U, Wahlberg N (2009) Phylogeny and biogeography of Coenonympha
677 butterflies (Nymphalidae: Satyrinae)–patterns of colonization in the Holarctic.
678 *Systematic Entomology*, **34**, 315–323.

679 Kunte K, Shea C, Aardema ML *et al.* (2011) Sex Chromosome Mosaicism and Hybrid
680 Speciation among Tiger Swallowtail Butterflies. *PLoS Genet*, **7**, e1002274.

681 Lafranchis T (2000) Les Papillons de jour de France, Belgique et Luxembourg et leurs
682 chenilles. *Collection Parthénopé, éditions Biotopé*.

683 Larsen PA, Marchan-Rivadeneira MR, Baker RJ (2010) Natural hybridization generates
684 mammalian lineage with species characteristics. *PNAS*, **107**, 11447–11452.

685 Leaché AD, McGuire JA (2006) Phylogenetic relationships of horned lizards (Phrynosoma)
686 based on nuclear and mitochondrial data: Evidence for a misleading mitochondrial
687 gene tree. *Molecular Phylogenetics and Evolution*, **39**, 628–644.

688 Liu B, Abbott RJ, Lu Z, Tian B, Liu J (2014) Diploid hybrid origin of *Ostryopsis intermedia*
689 (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change.
690 *Molecular Ecology*, **23**, 3013–3027.

691 Lukhtanov VA, Shapoval NA, Anokhin BA, Saifitdinova AF, Kuznetsova VG (2015)
692 Homoploid hybrid speciation and genome evolution via chromosome sorting.
693 *Proceedings of the Royal Society of London B: Biological Sciences*, **282**, 20150157.

694 Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*,
695 **20**, 229–237.

696 Mallet J (2007) Hybrid speciation. *Nature*, **446**, 279–283.

697 Mavárez J, Linares M (2008) Homoploid hybrid speciation in animals. *Molecular Ecology*,
698 **17**, 4181–4185.

699 Mavárez J, Salazar CA, Bermingham E *et al.* (2006) Speciation by hybridization in
700 *Heliconius* butterflies. *Nature*, **441**, 868–871.

701 Mayr, E (1963) Animal Species and Evolution. *Harvard University Press*

702 Meng C, Kubatko LS (2009) Detecting hybrid speciation in the presence of incomplete
703 lineage sorting using gene tree incongruence: A model. *Theoretical Population*
704 *Biology*, **75**, 35–45.

705 Mérot C, Mavárez J, Evin A *et al.* (2013) Genetic differentiation without mimicry shift in a
706 pair of hybridizing *Heliconius* species (Lepidoptera: Nymphalidae). *Biological*
707 *Journal of the Linnean Society*, **109**, 830–847.

708 Mouret V, Guillaumet A, Cheylan M *et al.* (2011) The legacy of ice ages in mountain species:
709 post-glacial colonization of mountain tops rather than current range fragmentation
710 determines mitochondrial genetic diversity in an endemic Pyrenean rock lizard.
711 *Journal of Biogeography*, **38**, 1717–1731.

712 Nice CC, Gompert Z, Fordyce JA *et al.* (2013) Hybrid Speciation and Independent Evolution
713 in Lineages of Alpine Butterflies. *Evolution*, **67**, 1055–1068.

714 Nolte AW, Freyhof J, Stemshorn KC, Tautz D (2005) An invasive lineage of sculpins, *Cottus*

715 sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from
 716 hybridization between old phylogeographic groups. *Proceedings of the Royal Society*
 717 *of London B: Biological Sciences*, **272**, 2379–2387.

718 Nolte AW, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends in Genetics*,
 719 **26**, 54–58.

720 Paradis E, Blomberg S, Bolker B *et al.* (2014) *ape: Analyses of Phylogenetics and Evolution*.
 721 Available at <http://cran.r-project.org/web/packages/ape/index.html>.

722 Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive Introgression across Species
 723 Boundaries in *Heliconius* Butterflies. *PLoS Genet*, **8**, e1002752.

724 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq:
 725 An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and
 726 Non-Model Species. *PLoS ONE*, **7**, e37135.

727 Porter AH, Schneider RW, Price BA (1994) Wing pattern and allozyme relationships in the
 728 *Coenonympha arcania* group, emphasising the *C. gardetta-darwiniana* contact area at
 729 Bellwald, Switzerland (Lepidoptera, Satyridae). *Nota Lepidopterologica*, **17**, 155–174.

730 Rhymer JM, Simberloff D (1996) Extinction by Hybridization and Introgression. *Annual*
 731 *Review of Ecology and Systematics*, **27**, 83–109.

732 Rieseberg LH (1997) Hybrid origins of plant species. *Annual review of Ecology and*
 733 *Systematics*, **28**, 359–389.

734 Roos C, Zinner D, Kubatko LS *et al.* (2011) Nuclear versus mitochondrial DNA: evidence for
 735 hybridization in colobine monkeys. *BMC Evolutionary Biology*, **11**, 77.

736 Rubin BER, Ree RH, Moreau CS (2012) Inferring Phylogenies from RAD Sequence Data.
 737 *PLoS ONE*, **7**, e33394.

738 Salazar C, Baxter SW, Pardo-Diaz C *et al.* (2010) Genetic Evidence for Hybrid Trait

739 Speciation in *Heliconius* Butterflies. *PLoS Genet*, **6**, e1000930.

740 Schmitt T, Besold J (2010) Upslope movements and large scale expansions: the taxonomy
741 and biogeography of the *Coenonympha arcania* – *C. darwiniana* – *C. gardetta* butterfly
742 species complex. *Zoological Journal of the Linnean Society*, **159**, 890–904.

743 Schmitt T, Müller P (2007) Limited hybridization along a large contact zone between two
744 genetic lineages of the butterfly *Erebia medusa* (Satyrinae, Lepidoptera) in Central
745 Europe. *Journal of Zoological Systematics and Evolutionary Research*, **45**, 39–46.

746 Schumer M, Rosenthal GG, Andolfatto P (2014) How Common Is Homoploid Hybrid
747 Speciation? *Evolution*, **68**, 1553–1560.

748 Schwarz D, Matta BM, Shakir-Botteri NL, McPherson BA (2005) Host shift to an invasive
749 plant triggers rapid animal hybrid speciation. *Nature*, **436**, 546–549.

750 Schwenk K, Brede N, Streit B (2008) Introduction. Extent, processes and evolutionary impact
751 of interspecific hybridization in animals. *Philosophical Transactions of the Royal
752 Society of London B: Biological Sciences*, **363**, 2805–2811.

753 Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*,
754 **19**, 198–207.

755 Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive Introgression of Anticoagulant
756 Rodent Poison Resistance by Hybridization between Old World Mice. *Current
757 Biology*, **21**, 1296–1301.

758 Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis
759 of Large Phylogenies. *Bioinformatics*, btu033.

760 Sun Y, Abbott RJ, Li L *et al.* (2014) Evolutionary history of Purple cone spruce (*Picea
761 purpurea*) in the Qinghai–Tibet Plateau: homoploid hybrid origin and Pleistocene
762 expansion. *Molecular Ecology*, **23**, 343–359.

763 Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography
 764 and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
 765 Thuiller W, Lavorel S, Midgley G, Lavergne S, Rebelo T (2004) Relating plant traits and
 766 species distributions along bioclimatic gradients for 88 leucadendron taxa. *Ecology*,
 767 **85**, 1688–1699.
 768 Tolman T, and Lewington R (2008) Guide des papillons d'Europe et d'Afrique du Nord.
 769 *Delachaux et Niestlé*.
 770 Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal
 771 in whitefish adaptive radiations. *Nature*, **482**, 357–362.
 772 Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide
 773 unprecedented resolution of species boundaries and relationships in the Lake Victoria
 774 cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
 775 Wiemers M (1998) *Coenonympha darwiniana* - a hybrid taxon? New insights through
 776 allozyme electrophoresis (Lepidoptera, Nymphalidae, Satyrinae).
 777 Yakimowski SB, Rieseberg LH (2014) The role of homoploid hybridization in evolution: A
 778 century of studies synthesizing genetics and ecology. *American Journal of Botany*.

779 **Data Accessibility:**

- 780 - De-multiplexed ddRAD sequencing data will be available at European Nucleotide
781 Archive (ENA) browser with the accession number “PRJEB10005”.
- 782 - Morphologic data will be available at Dryad Digital Repository in the form of the
783 coordinates of the 40 morphometric landmarks for each used individual.
- 784 - Digitalized pictures of the butterfly wings are freely available upon request from
785 T. Capblancq.
- 786 - A table with occurrence locations used for climatic niche comparisons will be
787 available at Dryad Digital Repository.
- 788 - Scripts are freely available upon request from T. Capblancq.

789

790 **Author contributions**

791 J. Mavarez, L. Després and T. Capblancq designed the study. T. C. and D. Rioux performed
792 ddRAD experiments. T. C. performed bioinformatics treatment of the resulted
793 sequences, as well as the genetic, morphologic and climatic analysis. T. C. wrote the
794 manuscript, and all authors contributed substantially to revisions.

Figure captions:

Figure 1: Map of the European Alps with the sampling locations for the three taxa in the complex: *Coenonympha arcania* (red squares), *C. gardetta* (blue circles) and *C. darwiniana* (green triangles). Right: picture of *C. arcania* (bottom) and *C. gardetta* (top) with the putative hybrid species (*C. darwiniana*) in the middle. (Pictures from <http://www.digitale-naturfotos.de> and <http://www.lepiforum.de>)

Figure 2: The 5 scenarios tested in the ABC analysis with model involving only cladogenesis events in the top and models involving hybridization in the bottom. A corresponds to the ancestral population, P1 to the current *C. arcania* population, P2 to *C. gardetta* population, P3 to French *C. darwiniana* population and P4 to Swiss *C. darwiniana*.

Figure 3: Results of the ABC procedure with the posterior probabilities of the different tested scenarios (top left), the retained scenario (bottom left, A corresponds to the ancestral population, P1 to *C. arcania* population, P2 to *C. gardetta* population, HP to an ancestral hybrid population, P3 to French *C. darwiniana* and P4 to Swiss *C. darwiniana*) and the posterior distribution of parameters corresponding to this scenario (right).

Figure 4: Principal Component Analysis obtained from the SNPs dataset. The first three PCs are represented with the percentage of information kept on each of them. The four main groups are represented by different symbols and colors: *C. arcania* (♦), *C. gardetta* (■), French *C. darwiniana* (▲) and Swiss *C. darwiniana* (●). Ellipses contain about 2/3 of the individuals for each group.

Figure 5: Individual wing shape and patterns discrimination (A) and relative wing size differences (B) with four groups: *C. arcania* (♦), *C. gardetta* (■), French *C. darwiniana* (▲) and Swiss *C. darwiniana* (●). The centroid size (relative wing size index) calculated during morphometric analysis shows significant differences among the four

studied groups and a Tuckey post hoc test highlighted size difference between each pair of population. A MANOVA realized on PCs scores after the Procrustes superimposition shows high differences among populations ($p\text{-value} < 2e\text{-}16$) and only a discriminant analysis performed on these PCs scores is represented in the figure. And (C) is a dendrogram showing morphological distances between groups.

Figure 6: Climatic niche separation in the OMI ordination analysis. The first axis corresponds to most of the precipitation/temperature gradients (see section Material and Method for more details). The second axis corresponds to mean temperature during the wettest (top of the graph) and driest (bottom) period of the year.

837 **Supporting information**

838 The following supplementary material is available for this article:

839

840 **Table S1: Table of the different studied populations** with geographic coordinates,
841 country, species and number of used individuals for the genetic and the morphological
842 analysis.

843

844 **Table S2: Table of the different prior parameter ranges and distributions included**
845 **in the ABC analysis.** Time scale corresponds to a number of generations (equal to the
846 number of years in our case).

847

848 **Figure S3: Landmarks used for geometric morphometrics,** shown on a ventral view
849 of *C. arcania* wings.