



**HAL**  
open science

## 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<sup>1,2\*</sup>, Laurence Després<sup>1,2</sup>, Delphine Rioux<sup>1,2</sup> and Jesús Mavárez<sup>1,2</sup>

<sup>1</sup> *Université Grenoble Alpes, LECA, F-38000 Grenoble, France*

<sup>2</sup> *CNRS, LECA, F-38000 Grenoble, France*

Date: July 29<sup>th</sup>, 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

1 **ABSTRACT:**

2

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

19

20 **Keywords:**

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

23           **INTRODUCTION:**

24           Hybridization between well-established species is known to be frequent in  
25 nature (Rieseberg 1997, Mallet 2007, Mavarez and Linares 2008, Schwenk *et al.* 2008,  
26 Descimon and Mallet 2010, Abbott *et al.* 2013), with around 10% of animal species and  
27 25% of plants involved in interspecific crosses according to Mallet 2005. However, and  
28 particularly among zoologists, this process has for long been considered to be either a  
29 genetic dead end leading mainly to sterile hybrids or to have no major effect in the  
30 divergence of species, even when fertile hybrids are produced (Mayr 1963). Indeed, it is  
31 generally accepted that gene flow through hybridization between diverging lineages can  
32 delay the achievement of reproductive isolation by homogenizing the gene pools,  
33 working against the speciation process (Elgvin *et al.* 2011, Yakimowski and Rieseberg  
34 2014). These views have been challenged during the last decades and interspecific  
35 hybridization is now accepted as a major process with a profound impact on several  
36 important evolutionary phenomena such as adaptation (Gompert *et al.* 2008, Kim *et al.*  
37 2008, Song *et al.* 2011, The Heliconius Genome Consortium 2012, Pardo-Diaz *et al.*  
38 2012), extinction (Rhymer and Simberloff 1996, Vonlanthen *et al.* 2012, Hohenlohe *et al.*  
39 2010) and diversification/speciation (Mallet 2007, Mavarez and Linares 2008, Nolte and  
40 Tautz 2010, Nice *et al.* 2013, Yakimowski and Riesberg 2014). In particular, there is a  
41 recent interest in studying the contribution of hybridization to speciation through the  
42 formation of new hybrid taxa. This has led to the identification of hybrid species in  
43 diverse animal groups including insects (Mavarez *et al.* 2006, Salazar *et al.* 2010, Kunte  
44 *et al.* 2011), birds (Brelsford *et al.* 2011, Hermansen *et al.* 2011), fishes (Nolte *et al.*  
45 2005) and mammals (Amaral *et al.* 2014). These studies, mostly conducted during the

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

58         Hybrid species appear to be quite rare in nature, but the expanding literature on  
59 this topic is beginning to provide valuable information regarding the patterns associated  
60 with this phenomenon (Seehausen 2004, Mallet 2007, Mavarez and Linares 2008,  
61 Abbott *et al.* 2013). Some studies have highlighted a primordial role of intermediate  
62 morphology in the emergence of the newborn taxon (Mavarez *et al.* 2006, Salazar *et al.*  
63 2010), while others have focused more on the crucial ecological adaptations brought by  
64 the genetic hybrid combination (Nolte *et al.* 2005, Schwarz *et al.* 2005, Gompert *et al.*  
65 2006, Kunte *et al.* 2011) or the intrinsic genomic mechanisms which have allowed the  
66 isolation of the hybrid lineage from parental species (Lukhtanov *et al.* 2015). Taken  
67 together, these studies suggest that even if population genetics provide definitely the  
68 best tools to identify hybrid species, multiple lines of evidence, including morphology,

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

77         Here we examine the role of hybridization in the evolutionary history of three  
78 European butterflies in the genus *Coenonympha* (Nymphalidae: Satyrinae). Two are well  
79 differentiated species: a) the Pearly Heath (*C. arcania*, Linné 1761), a common lowland  
80 butterfly present from Spain to Russia and b) the Alpine Heath (*C. gardetta*, Prunner  
81 1798) a species living above 1500m with a disjoint distribution including mainly the  
82 Alps, but also the French massif central, Bosnia and Herzegovina, and Albania (Tolman  
83 and Lewington 2009). These two species diverged 1.5 to 4 million years ago according  
84 to a phylogeny of the genus *Coenonympha* based on mitochondrial *COI* sequences  
85 (Kodandaramaiah and Wahlberg 2009). The third taxon, the Darwin's Heath (*C.*  
86 *darwiniana*) shows an intermediate morphology and completely replaces *C. gardetta* in  
87 two distinct and narrow areas: the Valais and Tessin regions in Switzerland and the  
88 South West of the French Alps. Limited contact zones between pairs of the three taxa  
89 have been reported (Porter *et al.* 1995) but on-going gene flow between them is  
90 unknown. The taxonomic status of *C. darwiniana* has been hotly debated and specialists  
91 have variously classified it as a subspecies of *C. arcania*, a subspecies of *C. gardetta* or as  
92 hybrid between them (Porter *et al.* 1994, Wiemers *et al.* 1998, Schmitt and Besold 2010,

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

## 109 **MATERIAL AND METHODS:**

### 110 *Butterfly sampling*

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

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

#### 119 *Genetic data acquisition*

120 DNA was extracted from the complete thorax of each individual using the DNeasy  
121 Blood and Tissue Kit (QIAGEN, Germany) according to the manufacturer's instructions  
122 and stored at -20°C. A double-digested RAD (Restriction site Associated DNA)  
123 experiment was conducted on 174 individuals using a modified version of the protocol  
124 described in Peterson *et al.* (2012). In short, 300ng of DNA template from each  
125 individual were double-digested with 10 units each of *Sbf*I-HF and *Msp*I (New England  
126 Biolabs Inc.) at 37°C during one hour in a final volume of 34µl using the CutSmart buffer  
127 provided with the enzymes. Digestion was further continued together with ligation of P1  
128 and P2 adapters (see Peterson *et al.* 2012) by adding 10 units of T4 DNA ligase (New  
129 England Biolabs Inc.), adapters P1 and P2 in 10-fold excess (compared to the estimated  
130 number of restriction fragments) and 1µl of 10mM ribo-ATP (New England Biolabs Inc.)  
131 in each sample. This simultaneous digestion-ligation reaction was performed on a  
132 thermocycler using a succession of 60 cycles of 2 min at 37°C for digestion and 2 min at  
133 16°C for ligation, after which the enzymes were heat inactivated at 65°C for 10 minutes.  
134 An equal volume mixture of all the digested-ligated fragments were purified with  
135 Agencourt AMPure XP beads (Beckman Coulter, France). Fragments were size-selected  
136 between 250 and 500bp by gel extraction (agarose gel: 1.6%) and purified with  
137 QIAquick Gel Extraction Kit (Qiagen, Germany). The ddRAD library obtained was  
138 amplified using the following PCR mix: a volume of 20µl with 1µl of DNA template, 10



139 mM of dNTPs, 10 $\mu$ M of each PCR probe (Paterson *et al.* 2012) and 2U/ $\mu$ l of *Taq* Phusion-  
140 HF (New England Biolabs Inc.); and the following PCR cycles: an initial denaturation at  
141 98°C for 10 minutes; 15 cycles of 98°C for 10 sec, 66°C for 30 sec and 72°C for 1 min;  
142 followed by a final extension period at 72°C for 10 min. The amplified ddRAD library  
143 was purified with QIAGEN MinElute PCR Purification Kit (Qiagen, Germany) and  
144 sequenced on an Illumina Hi-Seq 2000 (4/10 lane, paired-end 2 x 100bp, Fasteris SA,  
145 Switzerland).

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

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

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

162 from all the individuals' loci data sets was then built using the *cstacks* function in order  
163 to find all the alleles in the libraries and to produce a catalog of consensus alleles for  
164 each locus. A maximum of 5 mismatches was allowed for considering two individual tags  
165 as the same locus and to merge them in the catalogue (-n). The sets of loci constructed  
166 within each individual read pool were then searched against the newly produced catalog  
167 (*sstacks* function) and a SNP dataset was produced with the genotype of each individual  
168 for every polymorphic position (*populations* function). Only tags present in more than  
169 60% of the sampling were used to find the SNPs in order to avoid an excess of missing  
170 data. Finally, SNPs with a frequency lower than 5% were removed from the data set and  
171 only one polymorphic site was kept for each RAD-tag. This genetic matrix was then used  
172 to test different models of evolution.

173         Several scenarios have been proposed for the evolutionary history of these three  
174 *Coenonympha* species and especially for the origin of *C. darwiniana*. To compare them  
175 and find the most plausible one we used the SNP dataset with the Approximate Bayesian  
176 Computation (ABC) approach implemented in the software DIYABC (Cornuet *et al.*  
177 2014). In short, a large number of simulated data sets were produced under each tested  
178 scenario by sampling parameter values into predefined prior distributions and an  
179 estimation of each scenario posterior probability was calculated as a function of the  
180 similarity between several summary statistics of simulated data sets and observed data  
181 using a logistic regression procedure. These posterior probabilities enable a ranking of  
182 scenarios and allow for the identification of the most probable evolutionary history. We  
183 only considered five evolutionary scenarios (Fig 2), based on previous published studies  
184 dealing with the evolution of the complex (Porter *et al.* 1995, Wiemers *et al.* 1998 and

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

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

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

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

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

## 226 *Multivariate analysis*

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

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

### 232 *Geometric morphometry analysis*

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

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

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

#### 261 *Climatic niche comparisons*

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

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

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

## 289 **RESULTS:**

### 290 *Genetic evidence of hybridization*

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

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

307         It is relevant to observe that scenario 4 is the one with the best fit (40% of the  
308 summary statistics calculated on datasets simulated with scenario 4 are very close to the  
309 summary statistics calculated with the real genetic dataset) in comparison with the  
310 other four scenarios tested (only 19 to 33%). Our model produces simulations slightly  
311 distant from the observed genetic dataset in the summary statistics space, but given the  
312 simplicity of designed scenarios, it is not surprising that they do not perfectly match the  
313 real history of our populations and do not produce simulations in complete concordance  
314 with observed data. Moreover, type I (11.4%) and type II (0.05%) errors were very low  
315 in comparison with similar studies (Barres *et al.* 2012, Cornuet *et al.* 2010). The scenario  
316 4 was thus greatly distinguishable from the others. Taken together these scores in post  
317 ABC procedure verifications strongly support the evolutionary scenario 4, *i.e.* a  
318 hybridization event, at the origin of *C. darwiniana*.

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



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

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

343 *Morphological proximities among the complex*

344 A total of 510 specimens had both right fore- and right hindwings in good  
345 conditions and were included in the morphological analysis. A major result of the  
346 morphometric analysis is the good concordance they show with the genetic clustering  
347 analysis (Fig 5A). We found highly significant differences in size (ANOVA: p-value < 2e-  
348 16, df = 3, F-value = 172.2) and shape (MANOVA: p-value < 2e-16, df = 3, approx. F-value  
349 = 14.5) among the four previously established genetic clusters. For these two traits the  
350 strongest differentiation is visible between *C. arcania* and *C. gardetta* individuals with  
351 almost no overlap in wing centroid size values (Fig 5B) and in scores of the discriminant  
352 analysis first axis (Fig 5A). This first axis of the CVA combines the majority of the  
353 discriminant variability (80%) and separates clearly the four groups, with both *C.*  
354 *darwiniana* populations taking intermediate values between the two putative parental  
355 species (Fig 5A-C). Other analyses of landmarks corresponding only to either venation,  
356 the shape of the hindwing white band or only the positions of hindwing eyespots  
357 confirm this intermediacy (not shown). Our results robustly support the intermediate  
358 position of *C. darwiniana* in regards to the putative parental species, not only for easily  
359 visible characters in the phenotypes such as the eyespots' color or their alignment, as  
360 previously highlighted by some experts, but also for most of the wing structures.

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

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

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

### 373 *Climatic niches analysis*

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

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

#### 394 **DISCUSSION:**

##### 395 *Hybrid species hypothesis for Coenonympha darwiniana*

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

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

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

434 **Genetics:** PCA analysis provides an idea of genetic clustering without making any  
435 assumptions and supports the hybrid origin scenario. The two *C. darwiniana* lineages  
436 have intermediate scores between *C. arcania* and *C. gardetta* along the first PC which  
437 reflects the highest genetic structure within the complex. Although not a formal proof of  
438 hybrid origin, this result greatly concords with that hypothesis. The successive  
439 differentiation shown by the first three PCs completely correspond to the successive  
440 events of the hybrid species scenario (parental differentiation, hybridization event and  
441 hybrid lineages differentiation, see Figure 3). As stated above, the first PC shows the  
442 ancestral divergence of the lowland and alpine species with intermediacy of the two  
443 lineages of *C. darwiniana*. The second axis corresponds to the distinction of the hybrid  
444 lineages on one side and the parental species on the other side, showing the intrinsic  
445 genetic variability of these taxa and their history since the speciation event. Finally, PC3  
446 discriminates the two *C. darwiniana* populations, consistent with the geographic  
447 location of the individuals. This result increases our confidence in our conclusions and  
448 adds support to the view that ddRAD sequencing can be a powerful technique for  
449 genetic clustering at a fine scale (Wagner *et al.* 2012, Rubin *et al.* 2012, Cruaud *et al.*  
450 2014).

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

457 shown in North American *Lycaeides* butterflies (Gompert *et al.* 2009, Nice *et al.* 2013).  
458 But the ABC modeling clearly rejected this possibility and converged toward a unique  
459 and common origin for the two *C. darwiniana* taxa. Given the recent origin of the taxon  
460 (< 20000 years), the divergence observed between them seems large to be due only to  
461 independent evolution after geographical splitting. A possible explanation would be a  
462 differential backcrossing rate with the parental species for each of the two *C. darwiniana*  
463 lineages. Under this hypothesis, independent parental genetic contributions posterior to  
464 the lineage splitting observed within *C. darwiniana* would explain the substantial  
465 separation. The genetic stability evident now in the populations would be the result of  
466 their repeated interactions with one or both parental species, possibly driven by  
467 different environmental pressures. However, this hypothetical scenario remains  
468 uncertain and needs to be tested. It could provide an interesting case of successive  
469 hybridization events leading to differential adaptations, and illustrate the tenuous  
470 equilibrium between backcrossing due to incomplete reproductive isolation of the  
471 newborn lineage and success of hybrid population establishment.

472 **Morphology:** Morphology can play an important role in species isolation for butterflies  
473 and other animals because of visual preferences associated with mating behaviors. This  
474 is one of the reasons for which it has been investigated for most animal hybrid species.  
475 In some cases the hybrid lineage shows extreme traits compared to the parental species  
476 (e.g. a sculpin lineage in Nolte *et al.* 2005 or the bat *Artibeus schwartzi* in Larsen *et al.*  
477 2010), with these extreme phenotypes being potentially advantageous in species fitness.  
478 But most hybrid species present an intermediate morphology (e.g. the hybrid butterfly  
479 *Heliconius heurippa* in Mavarez *et al.* 2006 and Salazar *et al.* 2010, the birds *Passer*

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

501 **Ecology:** Ecology is known to play a major role in hybrid species emergence (Mallet  
502 2007, Mavarez and Linares 2008, Abbott *et al.* 2013). The mix of different genetic



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

525           Our study provides strong evidence of hybridization between two distinct  
526 *Coenonympha* species in the composition of *C. darwiniana* and a substantial reproductive  
527 isolation as suggested by the genetic, morphologic and climatic clustering of the taxa.  
528 However, an important additional aspect needs to be evaluated: the evidence that  
529 reproductive isolation between the hybrid and parental lineages is a direct consequence  
530 of hybridization (Mavarez and Linares 2008, Nolte and Tautz 2010, Schumer *et al.*  
531 2014). It is difficult with our results to determine whether hybridization has been the  
532 key mechanism of the speciation process (as shown for example for a sculpin hybrid  
533 species in Nolte *et al.* 2005, or for a hybrid butterfly species in Mavarez *et al.* 2006) or  
534 whether a lineage already diverged from one of the parental species has undergone a  
535 strong introgression from the other. However, the current stability, homogeneity and  
536 specificities of *C. darwiniana* in its two geographic populations together with the  
537 complete exclusion of parental species in these regions suggests an important role of  
538 hybrid patterns in the adaptation and persistence of this taxon.

#### 539           *Structure of the complex and taxonomic considerations*

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

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

572 (2010) by giving species status to each of the two currently recognized sub-species of *C.*  
573 *darwiniana*. Their disjoint geographic ranges together with their differences in genetic  
574 composition, wing structure and patterns and, to a lesser extent, their climatic niches,  
575 seem to us as enough evidence to justify their species status. We propose to raise the  
576 names previously used for the subspecies (Schmitt and Besold 2010) to a species level:  
577 *Coenonympha darwiniana* for the Tessino population and *Coenonympha macromma* for  
578 the southwest Alps population.

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

584

585

**ACKNOWLEDGEMENTS:**

586

We are very grateful to Flavia entomologist association, Mercantour National Parc,

587

French National Natural History Museum and Asters conservatory for their help in data

588

collection. We would also like to thank Yann Baillet, Pascal Dupont, Marie-France Leccia,

589

Gregory Guichert and Eric Drouet for helping us collecting samples and giving us helpful

590

advises on the project. Finally we thank Luis Daniel Llambi, Eric Bazin and Frédéric

591

Boyer for comments and revision of the manuscript. This project was funded by the

592

“Conseil general de l’Isère” and a “CNRS INEE – APEGE” grant.

593

594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617

**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 Parthénope, éditions Biotope*.

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. garetta-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.

795 **Figure captions:**

796

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

802

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

808

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

815

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

821

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

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

831

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

836



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.