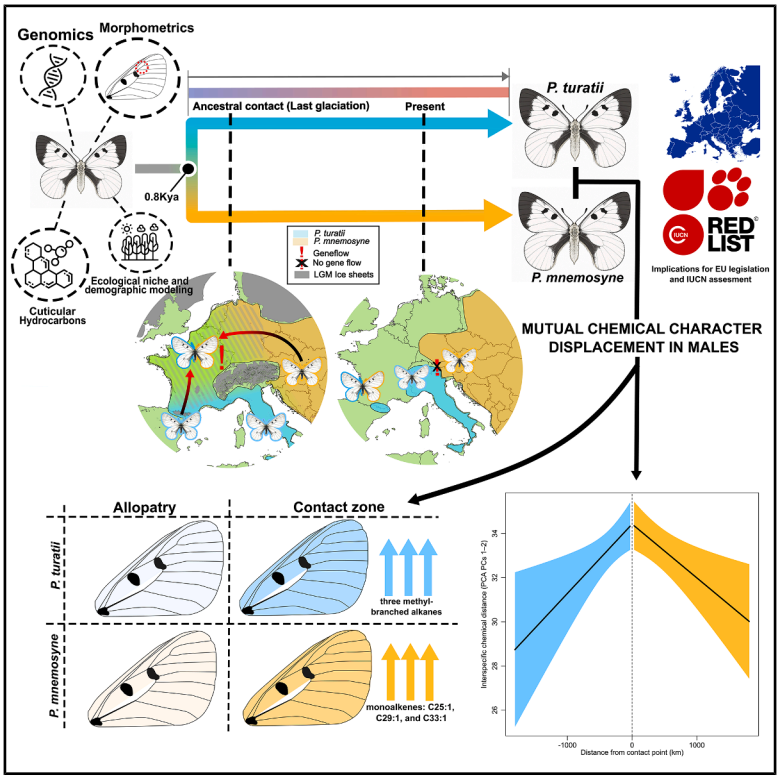
Report

Current Biology

Second-chance speciation in the Clouded Apollo butterfly

Graphical abstract



Highlights

- Genomics reveal that Europe's iconic Clouded Apollo butterfly includes two species
- An ancient contact with gene flow formed the admixed Pyrenean populations
- In the current contact, reproductive isolation is linked to male chemical divergence
- A case of recent chemical speciation with profound consequences for conservation

Authors

Josep Lancho-Silva, Loukia Spilani, Valéria Marques, ..., Dean C. Adams, Leonardo Dapporto, Roger Vila

Correspondence

josep.lancho@ibe.upf-csic.es (J.L.-S.), leonardo.dapporto@unifi.it (L.D.), roger.vila@csic.es (R.V.)

In brief

Lancho-Silva et al. reveal two successive contacts between butterfly populations that capture the surprisingly rapid emergence of reproductive isolation driven by selection for chemical differences, a mechanism rarely documented in action. Their work demonstrates that Europe's protected Clouded Apollo butterfly actually comprises two distinct species, with implications for conservation.

Lancho-Silva et al., 2026, Current Biology *36*, 1–9
August 3, 2026 © 2026 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cub.2026.06.052

CellPress

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

Current Biology

CellPress
OPEN ACCESS

Report

Second-chance speciation in the Clouded Apollo butterfly

Josep Lancho-Silva,[1,2,*] Loukia Spilani,[1,2] Valéria Marques,[1] Claudia Bruschini,[3] Vlad Dincă,[4] Marko Mutanen,[5] Markus Franzén,[6] Dean C. Adams,[7] Leonardo Dapporto,[3,*] and Roger Vila[1,8,*]

[1]Institut de Biologia Evolutiva, CSIC, Universitat Pompeu Fabra, Passeig Marítim de la Barceloneta 37, Barcelona 08003, Spain
[2]Faculty of Biology, University of Barcelona, Av. Diagonal 643, Barcelona 08028, Spain
[3]Dipartimento di Biologia, Università degli Studi di Firenze, Via Madonna del Piano 6, Sesto Fiorentino, Firenze 50019, Italy
[4]"Grigore Antipa" National Museum of Natural History, Şos. Kiseleff no. 1, Bucharest 011341, Romania
[5]Ecology and Genetics Research Unit, University of Oulu, Pentti Kaiteran katu 1, Oulu 90014, Northern Ostrobothnia, Finland
[6]IFM Biology, Conservation Ecology Group, Linköping University, Linköping 581 83, Sweden
[7]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, 251 Bessey Hall, Ames, IA 50011, USA
[8]Lead contact
*Correspondence: josep.lancho@ibe.upf-csic.es (J.L.-S.), leonardo.dapporto@unifi.it (L.D.), roger.vila@csic.es (R.V.)
https://doi.org/10.1016/j.cub.2026.06.052

SUMMARY

Contact zones, where genetic lineages meet, provide a unique opportunity to study speciation and reveal how reproductive barriers emerge.[1–3] However, species ranges are dynamic, contact zones may be ephemeral,[4,5] and reinforcement—selection against hybrids—is an elusive phenomenon.[6–8] We use an integrative approach—genomic, chemical, and morphological data coupled with ecological and demographic modeling—to evaluate lineage limits, reconstruct contact history, and infer reproductive isolation strength and mechanisms in the protected Clouded Apollo butterfly (*Parnassius mnemosyne*) in Europe.[9] The status of this taxon is highly debated: a potential cryptic species (*Parnassius turatii*) edging *P. mnemosyne sensu stricto* in the Eastern Alps has been proposed, yet evidence remains inconsistent in the absence of genomic data.[10–18] We reveal two genetic lineages that diverged about 0.8 million years ago and experienced two successive episodes of secondary contact. During the last glaciation, the lineages met in western Europe, which resulted in admixture and the formation of hybrid Pyrenean populations before reproductive isolation had evolved. In contrast, genomic data show no detectable gene flow at their present-day contact zone in the Eastern Alps, despite morphological and ecological similarity. Evidence of mutual chemical character displacement in males at this contact suggests bidirectional reinforcement in action. These findings portray a system in which lineages that once exchanged genes now re-encounter each other and reproductive isolation is generated—a second chance for speciation to succeed—and demonstrate that the Clouded Apollo comprises two cryptic species, with implications for EU legislation implementation and IUCN assessments.[9,19,20]

RESULTS

Genome-wide markers resolve two species with a sharp boundary at the contact zone

Analyses based on genomic data consistently separated *P. mnemosyne* and *P. turatii* into two well-supported genomic clusters. The phylogenetic analysis recovered two major clades corresponding to *P. mnemosyne* and *P. turatii*, with high support: one group was represented by samples from Spain, France, the Apennines (including Sicily), and the western/central Alps, while the remaining *P. mnemosyne* formed a distinct clade (Figure 1A1). Principal-component analysis (PCA) segregated individuals by species and geography, with PC1 (>18% variance) separating samples on either side of the contact zone and placing a small number of peripheral individuals intermediate, thus indicating a potential ancestral admixture outside the current contact (Figures 1C1 and 1C2). Bayesian ancestral clustering analysis cleanly separated the species with no admixture at the current contact zone, while the best supported was K = 3, which further subdivided *P. mnemosyne*; K = 4 identified the Pyrenean population as a discrete lineage (Figure 1A2). Geographic overlays of ancestry proportions show a strong phylogeographic structure across Europe (Figures 1B1 and 1B2). Importantly, the contact zone in the Eastern Alps demonstrates the strict genomic integrity of the two species.

Contemporary gene flow is absent at the Alpine contact zone, but ancient introgression into Pyrenean *P. turatii* is supported

Gene flow tests (D-statistics) across different population configurations revealed no detectable gene flow between contact-zone *P. mnemosyne* and *P. turatii*. In contrast, a significant gene flow between *P. mnemosyne* from the contact zone and Pyrenean *P. turatii* was detected (p = 0.0036; Figures 2A1 and

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

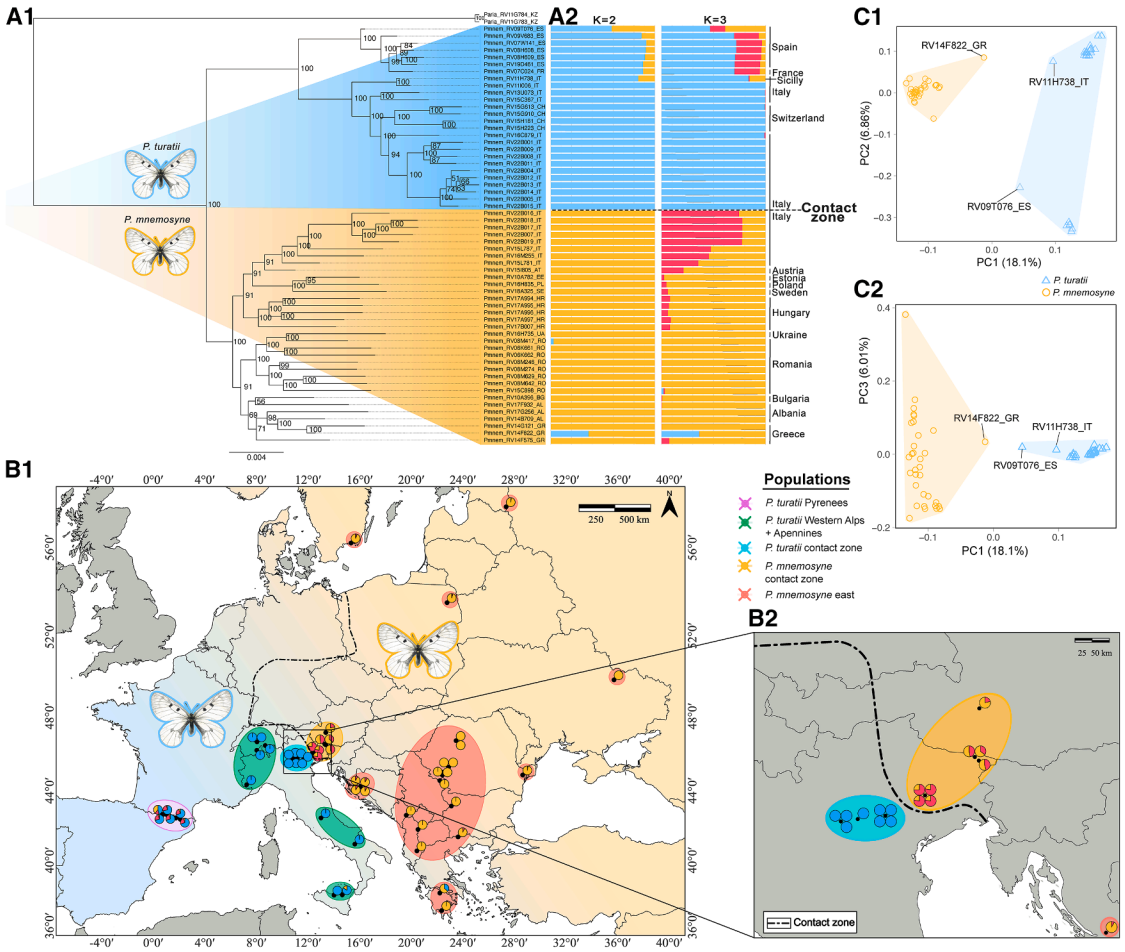


Figure 1. Phylogenetic and population structure analyses of the Clouded Apollo butterfly in Europe

(A1) Maximum likelihood phylogenetic tree showing two well-supported clades corresponding to the cryptic species *P. turatii* and *P. mnemosyne*. See also Table S1.

(A2) STRUCTURE results at K = 2 and K = 3, represented as bar plots showing individual ancestry proportions.

(B1 and B2) STRUCTURE results at K = 3 represented as pie charts overlaid on a geographic map, showing ancestry proportions per sampling locality; the zoom corresponds to the current contact zone in the Eastern Alps. Dashed line indicates the western limit of the distribution of *P. mnemosyne sensu stricto*, according to Bolotov et al.[13]

(C1 and C2) Principal-component analyses (PCAs) based on SNP data: PC1 vs. PC2 (top) and PC1 vs. PC3 (bottom); individuals with potentially admixed genomes, which are consistent with STRUCTURE results, are indicated.

2A2), despite the large geographic distance between these populations. Although surprising given their current distribution ranges, this result is consistent with STRUCTURE clustering signals of admixture in the Pyrenees. One individual from Sicily and one from the Peloponnese also showed signals of potential ancestral admixture, but this hypothesis could not be confirmed due to limited sample sizes. Across all individuals screened, no *Wolbachia* infection was detected, which argues against cytoplasmic incompatibility as a player in reproductive isolation.

Glacial-cycle dynamics created opportunities for secondary contact; demographic models quantify the divergence and asymmetric gene flow

Ecological niche models (ENMs) indicate some overlap of suitable areas for the two species. However, potentially suitable areas are markedly different in their continuity: *P. mnemosyne* shows a bigger and more continuous suitable area that extends to the east (current suitable area = 2.4 million km^2), whereas *P. turatii* is fragmented and mostly confined to a few Mediterranean and adjacent montane regions (0.7 million km^2; Figure 2B1, Current). During the last glacial maximum (LGM) (Figure 2B1, LGM; ~20 kya), both species were unable to occupy the completely glaciated northern Europe. We infer that, during this period, *P. mnemosyne* occupied large refugial belts in the Balkans/Black Sea-southern Carpathians-Caucasus foothills and, remarkably, western Europe to the Atlantic coast (4.5 million km^2). *P. turatii* potentially persisted in Iberia and the central Mediterranean, but the main suitable refugium was located in western Europe, from the Atlantic coast of France to the Alps (1.1 million km^2; Figure 2B1, LGM). The potential distributional overlap of the two species at the LGM (0.7 million km^2) was much more than that of today (0.4 million km^2) (Figure 2B2). The main area of overlap was recovered in western Europe, which is a

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

CellPress
OPEN ACCESS

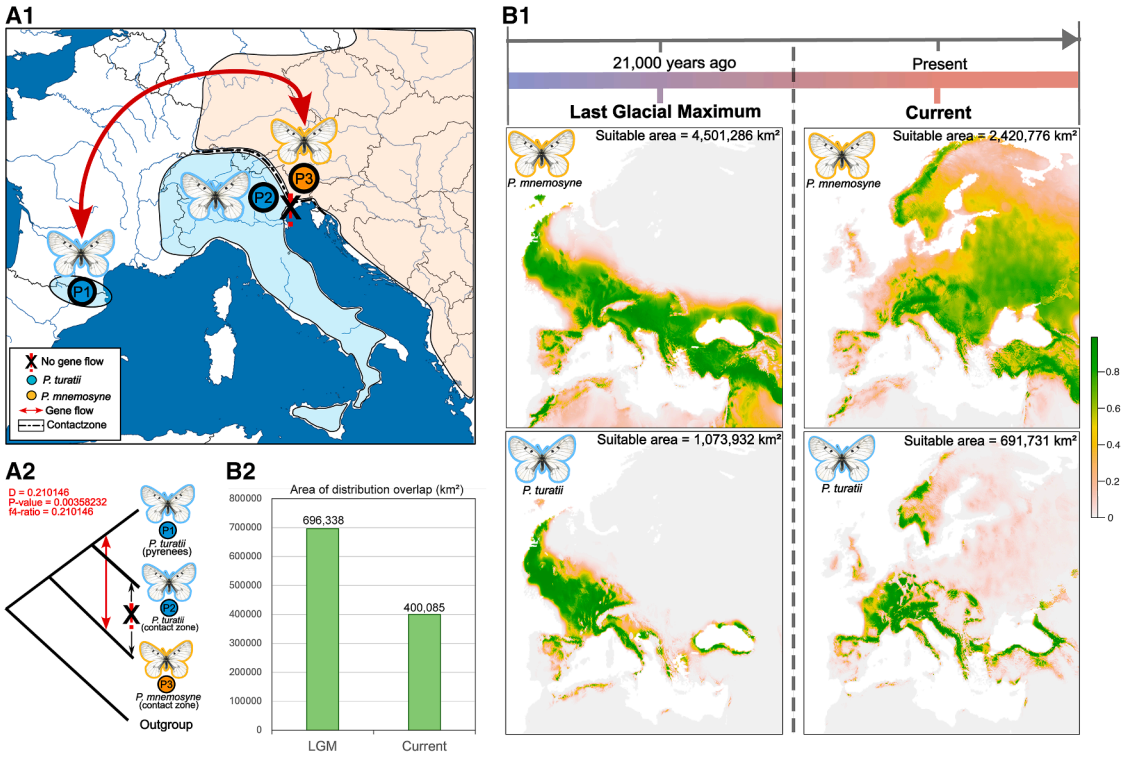


Figure 2. Ecological niche modeling and demographic history of *P. mnemosyne* and *P. turatii*

(A1) Map showing the geographic distributions of the populations used in one of the D-statistics gene-flow tests, with arrows indicating detected gene-flow events and crossed exclamation marks indicating no evidence of gene flow. See also Figure S2.

(A2) Schematic tree summarizing the relationships among the analyzed populations and the inferred interspecific gene-flow events. Note that additional tests using other populations were performed that confirmed this result. See also Figure S2.

(B1) Ecological niche model projections for *P. mnemosyne* and *P. turatii* during the last glacial maximum (LGM) and under current climatic conditions.

(B2) Estimated area of potential distribution overlap between the two species (in km^2) under LGM and present-day climatic scenarios.

main potential refugium for both lineages and the area where contact likely occurred during the last glaciation.

Demographic inference revealed a first split at 819 kya, with a subsequent secondary split within *P. turatii* at 154 kya, and unidirectional gene flow from *P. mnemosyne* into Pyrenean *P. turatii* (Figure 3). Alternative models invoking bidirectional exchange, hybrid origin, no gene flow, or alternative topologies showed a substantially worse fit to the data (Table S2). The quantitative spatiotemporal concordance between the ENM-inferred LGM overlap in western Europe and demographic modeling estimates support an ancient secondary contact event with asymmetric introgression in this region during the last glaciation, which has left a genetic footprint in the populations of the Pyrenees (Figures 2 and 3). The inferred effective population sizes of the main lineages are remarkably different, especially those of *P. mnemosyne* (N_e = 1.2 million individuals) and of the Pyrenean *P. turatii* (N_e = 77,000 individuals). Such a demographic imbalance could explain the asymmetry in gene flow, although this remains a speculative hypothesis and other factors may have played a role.

Wing scale morphology does not differ between species at the contact zone

Geometric morphometrics of forewing scales (Figure 4A)—a character previously suggested to have diagnostic value in this system[13]— revealed differences in shape correlating with species across their range: PCA captured 71.56% (PC1) and 16.66% (PC2) of variance, with allopatric *P. mnemosyne* skewed to negative PC1 and *P. turatii* (western Alps, Apennines, and Pyrenees) to positive PC1. Nevertheless, contact-zone individuals were intermediate regardless of their genetic lineage (Figure 4C). Statistical tests detected a significant group effect ($F = 12.07$, $Z = 6.4323$, $p = 0.001$; 19% variance explained). Pairwise tests among groups showed that allopatric *P. mnemosyne* differed from all *P. turatii* groups ($p \leq 0.003$), with Pyrenean *P. turatii* the most divergent ($p \leq 0.003$ vs. all others); differentiation between *P. mnemosyne* and *P. turatii* at the contact zone was marginal ($p = 0.054$) (Table S5). Deformation grids indicated narrower, elongated scales in *P. mnemosyne* and broader, more curved scales in *P. turatii* (Figure 4B), with the Pyrenean lineage showing the greatest lateral divergence.

Male cuticular hydrocarbons show maximum differentiation at the contact zone declining with distance

In total, 43 cuticular hydrocarbon (CHC) compounds were identified from the wings of *P. turatii* and *P. mnemosyne* males, comprising a mixture of straight-chain alkanes, methyl-branched alkanes, and monoalkenes (Table S6). Our chemical analyses (Figure 4A) revealed a separation between species overall,

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

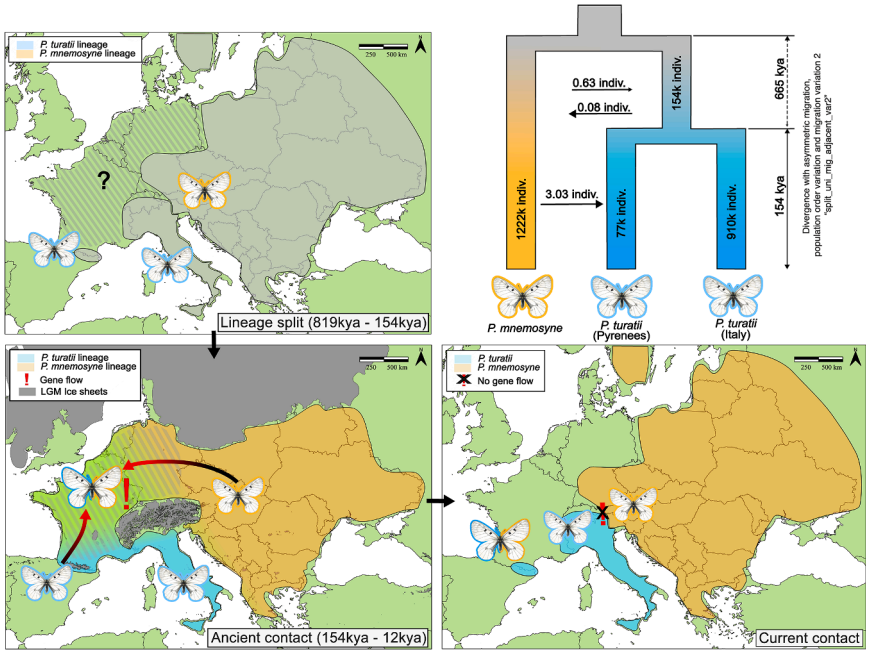


Figure 3. Demographic model and inferred evolutionary history of *P. mnemosyne* and *P. turatii*

The best-fit demographic model (top right) supports an initial evolutionary split (819 kya) followed by a more recent divergence within *P. turatii* (154 kya). Lineage relationships are shown, with arrows indicating the direction and relative magnitude of inferred gene flow. Note the substantial gene flow inferred from *P. mnemosyne* to the western branch of *P. turatii* (Pyrenees). Maps illustrate the three inferred phases: (1) lineage split, showing potential but unknown early allopatric ranges; (2) ancient contact (154–12 kya), when glacial-period expansions increased the geographic overlap and enabled secondary contact; and (3) present-day contact, where there is a sharp boundary and no current gene flow is detected.
See also Tables S2 and S3.

although individuals collected in the contact zone showed much clearer chemical differences: *P. turatii* and *P. mnemosyne* from the contact zone in the Eastern Alps occupied distinct regions of the first two PCA axes of chemical composition indicating the highest divergence between populations (Figure 4D). The compounds mostly associated with *P. turatii* are the three methyl-branched alkanes (3-MeC25, 13-MeC33, and 3Me-C36), while those associated with *P. mnemosyne* include specific isomers of three monoalkenes: C25:1 (isomers a and b), C29:1 (isomers a and b), and C33:1 (isomer b). Classification accuracy in a partial least square discriminant analysis for these contact populations was correspondingly high—96.0% for contact zone *P. turatii* and 100% for contact zone *P. mnemosyne*. In contrast, accuracy was much lower for populations farther from the contact zone (66.7% in non-contact *P. mnemosyne*, 63.6% in Pyrenean *P. turatii*, and 64.3% in western Alps-Italian peninsula *P. turatii*), although all were above 50%, which indicates some degree of differentiation.

As an exploratory analysis, we first examined the relationship between pairwise chemical and geographic distances across all individual comparisons (n = 2,850) with a multiple regression on distance matrices (MRM). The model (coefficient of determination R^2 = 0.0955, F statistic = 100.136, permutation-based significance *F.pval* = 0.0001) revealed the significant effects of geographic distance (regression coefficient estimate = 0.157, p = 0.0001), and pair type, comparing heterospecific vs. conspecific pairs (regression coefficient estimate = 10.924, p = 0.0001), as well as their interaction (regression coefficient estimate = −0.307, p = 0.0001). This significant interaction indicated opposite patterns within and between species: chemical divergence increased with geographic distance among conspecific pairs but decreased with geographic distance among heterospecific pairs. Thus, individuals from different species were most chemically differentiated when occurring in close geographic proximity.

To directly test this prediction, we then quantified male interspecific chemical divergence as a function of distance from the contact zone. Pairwise interspecific CHC distances (n = 2,600) declined significantly with increasing distance from the contact zone in both species (marginal R^2 = 0.032 and conditional R^2 = 0.216; Figures 4E and S1), consistent with symmetric character displacement. Species-specific models showed a similar decline in *P. turatii* (slope = −0.00314 ± 0.00116 SE, p = 0.0069) and *P. mnemosyne* (slope = −0.00244 ± 0.0009 SE, p = 0.0048), producing a mirror-like pattern in which chemical differentiation was strongest near the contact zone and progressively weakened with distance. Too few female individuals were available to perform similar analyses, but the preliminary results suggested low chemical interspecific differentiation in females.

DISCUSSION

The lack of gene flow at the contact zone in the Eastern Alps demonstrates that *P. mnemosyne* and *P. turatii* have completed speciation despite their recent divergence and apparent ecological similarity.

The concordance between gene-flow patterns, ancestral clustering analyses, demographic inference, and paleoclimate-based range reconstructions supports a coherent evolutionary scenario for this system. An initial lineage split about 0.8 mya separated the ancestors of *P. turatii* and *P. mnemosyne*, after which they likely diverged in allopatry for most of the time, although with some degree of gene flow. Because the divergence-time estimate relies on a mutation rate calibrated in *Heliconius melpomene*, it should be interpreted as approximate, but it nevertheless provides a useful temporal framework for comparing demographic history with climatic cycles. During the last glacial period (115–12 kya), both cold-adapted lineages were potentially more widely distributed and, remarkably, the ancestor of *P. mnemosyne* could have reached the Atlantic coast in western Europe. These glacial expansions brought the species into contact in western Europe, where they

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

CellPress
OPEN ACCESS

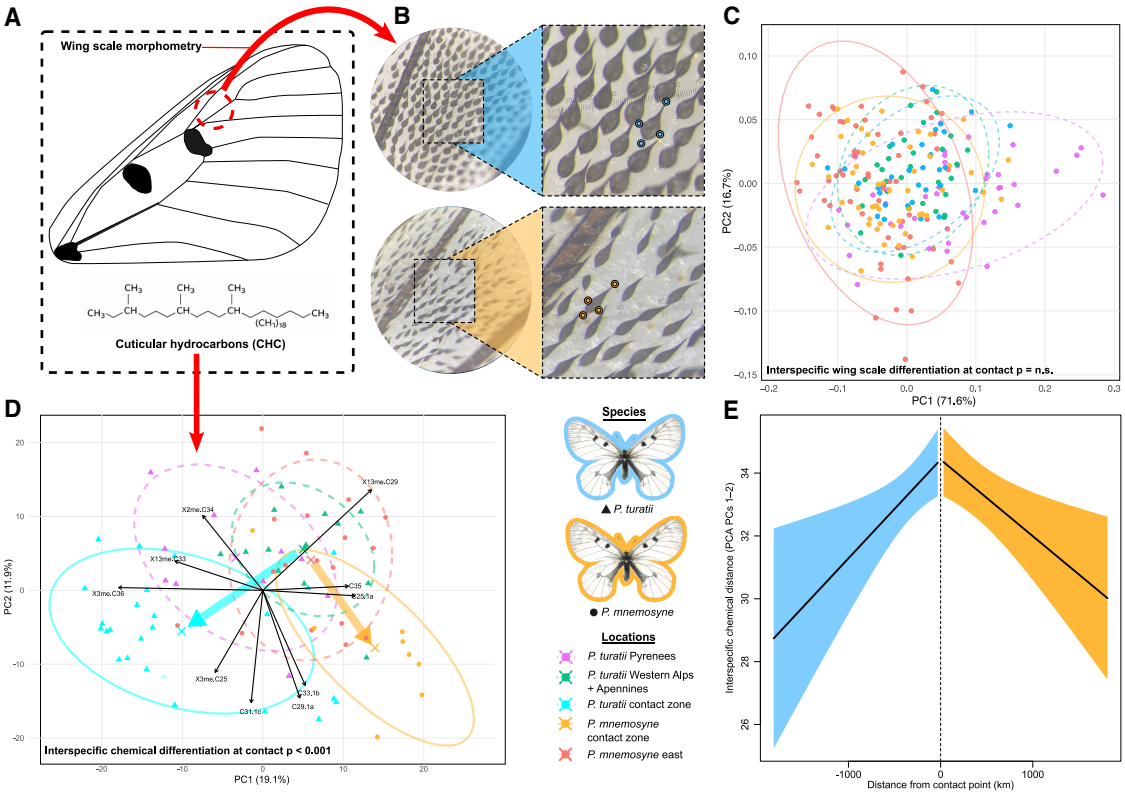


Figure 4. Morphological and chemical divergence between *P. mnemosyne* and *P. turatii*

(A) Schematic forewing showing the location of sampled scales and an example CHC compound.

(B) Representative scale morphology of *P. turatii* (blue) and *P. mnemosyne* (orange).

(C) PCA of scale shape (PC1 = 71.56%, PC2 = 16.66%). Note that interspecific wing scale differentiation at the contact zone is not significant. See also Tables S4 and S5.

(D) PCA of male CHC profiles reveals a partial overall separation but pronounced discrimination in the contact-zone populations, with key compounds (vectors) driving species chemical divergence. Colored arrows show chemical shifts by linking the centroids of non-contact populations to their conspecific contact-zone population. Interspecific chemical differentiation at the contact zone is highly significant. See also Table S6.

(E) Plot of male interspecific chemical distance vs. geographical distance to the center of the current ~40-km-wide contact zone in the Eastern Alps. Shaded areas represent 95% confidence intervals around the marginal fitted trends, excluding the random effect of focal individual.

See also Figure S1.

interbred and contributed to the admixed ancestry of today's Pyrenean populations. With the onset of interglacial warming, both lineages retreated to their current mountain refugia, thus reducing contact and allowing isolated populations to advance in the speciation process. Finally, upon their most recent secondary contact—represented by the current contact zone in the Eastern Alps—the lineages have apparently become reproductively isolated.

This system illustrates the view of speciation as a continuum, the final step of which usually takes place under secondary contact,[21–24] and provides a rare empirical example of what can be considered "second-chance speciation": a first failed opportunity for speciation during the last glaciation, in which gene flow dominated and the lineages fused, and a second successful one, in the current contact zone, where the clash between the forces of selection and admixture resulted in speciation. Earlier inferences of introgression between these lineages were based on a small number of highly polymorphic microsatellites,[18] and they can be reconciled with our genome-wide results by differences in marker resolution and sampling scale.

Genomic analyses provide us with the capacity not only to test for speciation in contemporary contact zones but also to infer ancestral events of sympatry and their outcome. It is likely that, given the repetitive Quaternary climatic oscillations, lineages had multiple contacts across recent evolutionary history, and each of them provided a chance for gene flow and reinforcement to act and for speciation to complete.[4,5,25–27] In the case of the Clouded Apollo cryptic species, we cannot rule out the existence of additional contact events in the past.[28] In fact, the demographic model suggests some degree of bidirectional gene flow across their history, and a few specimens in Sicily and the Peloponnesus show the signatures of potential ancestral admixture. How frequently contact events result in speciation is unknown, but we hypothesize that these cases represent a conspicuous minority, while lineage fusion may be a much more common outcome whose signal dilutes with time.

What factors could explain the fast emergence of a barrier to gene flow in the *Parnassius* Alpine contact zone? Their similar environmental requirements indicate that ecological divergence was probably not the primary driver of speciation. The

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

morphological divergence detected in wing scale shape, including clinal convergence near the contact zone, indicates phenotypic evolution along axes is only weakly coupled to genome-wide ancestry and was plausibly influenced by local environmental conditions such as montane humidity and UV exposure. Although previous taxonomic proposals mentioned this trait,[13] it is not a reliable diagnostic character, as the differences disappear in the contact zone, where the two species can most likely be misidentified. Genital mechanical isolation remains untested but is unlikely to be the main factor given that previous studies mention fairly minor genital differences.[13] The lack of *Wolbachia* infections argues against cytoplasmic incompatibility caused by these endosymbiotic bacteria.

A mechanism that plausibly may have finalized reproductive isolation is reinforcement generating interspecific chemical differentiation in males. CHC chemical divergence increased with geographic distance among conspecific individuals but decreased among heterospecific individuals, such that geographically close individuals from different species were the most chemically differentiated. Accordingly, we found mutual chemical character displacement in males localized at the contact zone without accompanying genomic admixture—an expected signature when selection against maladaptive hybridization strengthens premating cues. Similar potentially reinforcement-driven divergence has been described in *Agrodiaetus* butterflies, where sympatric sister species evolved stronger differences in male wing coloration following secondary contact between chromosomally differentiated lineages,[29,30] although the signaling modality in our system appears to be chemical rather than visual. Likewise, studies in *Heliconius* have shown that chemical signals can function as primary reproductive barriers even among co-mimetic taxa with similar wing patterns.[31] However, *Heliconius* also provides an interesting contrast, as speciation in this genus has often proceeded with persistent gene flow, whereas our genomic results suggest that gene flow has apparently ceased. Recent work on Alpine *Erebia* butterflies further suggests that CHC-mediated species recognition may be more widespread in European montane butterflies than previously appreciated, with CHC differentiation being higher between sympatric species pairs compared with allopatric ones.[32] In contrast to the *Erebia cassioides-E. tyndarus* hybrid zone, where reinforcement is not currently evident[33] the localized CHC divergence observed here may reflect ongoing reinforcement in sympatry, although the evidence remains correlative. This interpretation aligns with broader evidence that speciation in Lepidoptera can be completed through premating barriers even in the absence of large ecological niche shifts.[34,35]

Chemical communication is widespread in butterflies and is a well-established component of mate recognition, but to our knowledge, this is the first documented case in butterflies of bidirectional chemical character displacement, where both species show chemical divergence that reaches its maximum in sympatry and fades with distance to the contact. *Papilio polytes* use cuticular hydrocarbons for mate discrimination,[36–38] including alkene components, which are the main compounds causing the differentiation of *P. mnemosyne*. Behavioral mate choice tests, female-focused CHC sampling, and transcriptomic analyses of pheromone biosynthesis genes are logical next steps to establish causality.

Although our sampling was sufficient to recover the main phylogeographic, genomic, and CHC patterns, the limited number of individuals in some pooled populations and geographic regions means that low-level introgression, additional historical contact events, or finer-scale variation in CHC displacement cannot be excluded without broader sampling.

Because the Pyrenean populations are currently allopatric, their level of reproductive isolation cannot be determined. Although multiple hypotheses remain possible, we treat these populations as *P. turatii*, since only a small portion of their genome shows introgression from *P. mnemosyne*, and the phylogenetic analysis groups them with *P. turatii*. However, they are genetically distinct from populations in the Alps, Apennines, and Sicily and should be recognized at least as a subspecies. As the Pyrenees are the type locality for *turatii* Fruhstorfer, 1908—the name that has priority[14]—we propose two subspecies within *P. turatii*: *P. turatii turatii* Fruhstorfer, 1908 comb. nov. for the Pyrenean populations and *P. turatii parmenides* Fruhstorfer, 1908 comb. nov. (type locality, Maritime Alps) for the remaining populations studied.

Recognizing *P. mnemosyne* and *P. turatii* as distinct species, and dividing the latter into two subspecies, has important consequences for conservation. The redefinition of species boundaries results with two taxa with smaller effective distributions. Particularly *P. turatii* as it has highly localized and low-density populations in fragmented montane habitats that are vulnerable to climate warming.[39] For example, in the Catalan Pyrenees, where it is locally protected and constantly monitored, population density has dramatically declined (−95.3% from 2006 to 2025) and is apparently extinct in some regions.[40] Similarly, several Apennine populations seem to be on the brink of extinction.[19] Nevertheless, *P. mnemosyne* is also facing remarkable declines and extinctions in northern Europe.[41–43] Given the group's complex biogeographic history, a similar integrative analysis could be beneficial in Asia, a region that hosts far greater *Parnassius* diversity, with several lineages of unresolved taxonomic status. Combining genomic and phenotypic data will be essential to determine whether additional cryptic species occur there and to achieve a consistent, evolutionarily informed taxonomy.[16,44–46]

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to, and will be fulfilled by, the lead contact, Roger Vila (roger.vila@csic.es).

Materials availability

This study did not generate any new, unique reagents.

Data and code availability

- All genomic data associated with this paper are available in NCBI under BioProject: PRJNA1474983. Cuticular hydrocarbon data can be found at https://doi.org/10.5281/zenodo.20544798, and wing scale images can be found at https://doi.org/10.5281/zenodo.20542681.
- Custom code for cuticular hydrocarbon analysis can be found at https://doi.org/10.5281/zenodo.20544798, and custom code for wing scale geometric morphometry analysis can be found at https://doi.org/10.5281/zenodo.20542681.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

ACKNOWLEDGMENTS

We wish to thank M. Goia, Santi Viader, Juan Hernández-Roldán, Sylvain Cuvelier, Morten Molgaard, Raluca Vodă, Luca Casacci, E. Karolinskiy, Göran Sjöberg, and Joan Carles Hinojosa for providing specimens. We thank the authorities that issued the relevant authorizations to make this project possible. R.V. was supported by grant PID2022-139689NB-I00 (funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU) and J.L.-S. by the grant PREP2022-000487 (funded by MICIU/AEI/10.13039/501100011033 and ESF+).

AUTHOR CONTRIBUTIONS

Resources, J.L.-S., V.D., M.F., L.D., and R.V.; conceptualization, J.L.-S., L.D., and R.V.; formal analysis, J.L.-S., L.S., V.M., C.B., L.D., and D.C.A.; funding, M.M. and R.V.; investigation, J.L.-S., V.M., M.F., and V.D.; supervision, R.V.; writing – original draft, J.L.-S. and R.V.; visualization, J.L.-S., C.B., L.D., and D.C.A.; writing – review and editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Sampling
- METHOD DETAILS
  - ddRADseq library preparation
  - Wing scale imaging
  - CHC Extraction and GC-MS Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RADseq data processing
  - Population genomics analysis
  - Gene flow analysis
  - Ecological Niche Modeling
  - Demographic Modeling Inference
  - Wing scale morphometrics
  - Chemical discrimination and data preprocessing
  - *Wolbachia* infection

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2026.06.052.

Received: March 27, 2026
Revised: June 5, 2026
Accepted: June 19, 2026

REFERENCES

1. Butlin, R.K., Galindo, J., and Grahame, J.W. (2012). Genetics of ecological speciation. J. Evol. Biol. *25*, 2237–2254.
2. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates).
3. Jiggins, C.D., and Mallet, J. (2000). Bimodal hybrid zones and speciation. Trends Ecol. Evol. *15*, 250–255. https://doi.org/10.1016/S0169-5347(00)01873-5.
4. Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature *405*, 907–913. https://doi.org/10.1038/35016000.
5. Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. Front. Zool. *4*, 11. https://doi.org/10.1186/1742-9994-4-11.
6. Barton, N., and Bengtsson, B.O. (1986). The barrier to genetic exchange between hybridising populations. Heredity *57*, 357–376. https://doi.org/10.1038/hdy.1986.135.
7. Barton, N.H. (2020). On the completion of speciation. Philos. Trans. R. Soc. Lond., B Biol. Sci. *375*, 20190530. https://doi.org/10.1098/rstb.2019.0530.
8. Yukilevich, R. (2012). Asymmetrical patterns of speciation uniquely support reinforcement in Drosophila. Evolution *66*, 1430–1446. https://doi.org/10.1111/j.1558-5646.2011.01534.x.
9. European Union (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Communities *L 206*, 7–50. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1992L0043:20070101:en:PDF.
10. Gratton, P., Konopiński, M.K., and Sbordoni, V. (2008). Pleistocene evolutionary history of the Clouded Apollo (*Parnassius mnemosyne*): genetic signatures of climate cycles and a 'time-dependent' mitochondrial substitution rate. Mol. Ecol. *17*, 4248–4262. https://doi.org/10.1111/j.1365-294X.2008.03901.x.
11. Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M.L., García-Barros, E., Hebert, P.D.N., and Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. Sci. Rep. *5*, 12395. https://doi.org/10.1038/srep12395.
12. Dincă, V., Dapporto, L., Somervuo, P., Vodă, R., Cuvelier, S., Gascoigne-Pees, M., Huemer, P., Mutanen, M., Hebert, P.D.N., and Vila, R. (2021). High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. Commun. Biol. *4*, 315. https://doi.org/10.1038/s42003-021-01834-7.
13. Bolotov, I.N., Gofarov, M.Y., Gorbach, V.V., Kolosova, Y.S., Zheludkova, A.A., Kondakov, A.V., and Spitsyn, V.M. (2021). Parnassius nebrodensis: A threatened but neglected Apollo butterfly species from Southern Europe (Lepidoptera: Papilionidae). Ecol. Montenegrina *40*, 140–163. https://doi.org/10.37828/em.2021.40.13.
14. Cotton, A.M., Bolotov, I.N., Gofarov, M.Y., Gorbach, V.V., Kolosova, Y.S., Zheludkova, A.A., Kondakov, A.V., and Spitsyn, V.M. (2021). The correct name for the South Western European species recently separated from Parnassius mnemosyne (Linnaeus, 1758)(Lepidoptera: Papilionidae). Ecol. Montenegrina *43*, 56–58. https://doi.org/10.37828/em.2021.43.8.
15. Dapporto, L., Menchetti, M., Vodă, R., Corbella, C., Cuvelier, S., Djemadi, I., Gascoigne-Pees, M., Hinojosa, J.C., Lam, N.T., Serracanta, M., et al. (2022). The Atlas of mitochondrial genetic diversity for Western Palaearctic butterflies. Glob. Ecol. Biogeogr. *31*, 2184–2190. https://doi.org/10.1111/geb.13579.
16. Lukhtanov, V.A., and Zakharov, E.V. (2023). Taxonomic structure and wing pattern evolution in the *Parnassius mnemosyne* species complex (Lepidoptera, Papilionidae). Insects *14*, 942. https://doi.org/10.3390/insects14120942.
17. Höglund, J., Dias, G., Olsen, R.A., Soares, A., Bunikis, I., Talla, V., and Backström, N. (2024). A chromosome-level genome assembly and annotation for the clouded apollo butterfly (Parnassius Mnemosyne): A species of global conservation concern. Genome Biol. Evol. *16*, evae031. https://doi.org/10.1093/gbe/evae031.
18. Gratton, P., Todisco, V., Cesaroni, D., Sbordoni, V., and Nazari, V. (2025). Genetic data suggest gene flow within a narrow hybrid zone between two recently separated species in the genus Parnassius (Lepidoptera: Papilionidae). PLOS One *20*, e0321742. https://doi.org/10.1371/journal.pone.0321742.
19. Cini, A., Barbero, F., Bonelli, S., Bruschini, C., Casacci, L.P., Piazzini, S., Scalercio, S., and Dapporto, L. (2020). The decline of the charismatic Parnassius mnemosyne (L.) (Lepidoptera: Papilionidae) in a Central Italy

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

national park: a call for urgent actions. J. Insect Biodivers. *16*, 47–54. https://doi.org/10.12976/jib/2020.16.2.2.

20. Nadler, J., Benyamini, D., Bonelli, S., Comay, O., Dapporto, L., Karaçetin, E., Lukhtanov, V., López Munguira, M., Micevski, N., Settele, J., et al. (2021). Parnassius mnemosyne. IUCN Red List Threat. Species. https://doi.org/10.2305/IUCN.UK.2021-1.RLTS.T174210A122602056.en.
21. Lowry, D.B., and Gould, B.A. (2016). Speciation continuum. In Encyclopedia of Evolutionary Biology (Elsevier), pp. 159–165. https://doi.org/10.1016/B978-0-12-800049-6.00080-9.
22. Mallet, J., Beltrán, M., Neukirchen, W., and Linares, M. (2007). Natural hybridization in heliconiine butterflies: the species boundary as a continuum. BMC Evol. Biol. *7*, 28. https://doi.org/10.1186/1471-2148-7-28.
23. Stankowski, S., and Ravinet, M. (2021). Defining the speciation continuum. Evolution *75*, 1256–1273. https://doi.org/10.1111/evo.14215.
24. Shaw, K.L., and Mullen, S.P. (2014). Speciation continuum. J. Hered. *105*, 741–742. https://doi.org/10.1093/jhered/esu060.
25. Dapporto, L., Menchetti, M., Dincă, V., Talavera, G., Garcia-Berro, A., D'Ercole, J., Hebert, P.D., and Vila, R. (2024). The genetic legacy of the Quaternary ice ages for West Palearctic butterflies. Sci. Adv. *10*, eadm8596. https://doi.org/10.1126/sciadv.adm8596.
26. Marques, V., Hinojosa, J.C., Dapporto, L., Talavera, G., Stefanescu, C., Gutiérrez, D., and Vila, R. (2024). The opposed forces of differentiation and admixture across glacial cycles in the butterfly Aglais urticae. Mol. Ecol. *33*, e17304. https://doi.org/10.1111/mec.17304.
27. Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., and Cosson, J.F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. *7*, 453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x.
28. Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C.D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. *23*, 1817–1828. https://doi.org/10.1101/gr.159426.113.
29. Kandul, N.P., Lukhtanov, V.A., and Pierce, N.E. (2007). Karyotypic diversity and speciation in *Agrodiaetus* butterflies. Evolution *61*, 546–559. https://doi.org/10.1111/j.1558-5646.2007.00046.x.
30. Lukhtanov, V.A., Kandul, N.P., Plotkin, J.B., Dantchenko, A.V., Haig, D., and Pierce, N.E. (2005). Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies. Nature *436*, 385–389. https://doi.org/10.1038/nature03704.
31. González-Rojas, M.F., Darragh, K., Robles, J., Linares, M., Schulz, S., McMillan, W.O., Jiggins, C.D., Pardo-Diaz, C., and Salazar, C. (2020). Chemical signals act as the main reproductive barrier between sister and mimetic *Heliconius* butterflies. Proc. Biol. Sci. *287*, 20200587. https://doi.org/10.1098/rspb.2020.0587.
32. Kleckova, I., Clancy, M.V., Cornet, C., Alvarez, N., Matos-Maraví, P., and Lucek, K. (2025). Co-occurring sister taxa of mountain butterflies exhibit distinct cuticular hydrocarbon profiles. Ecol. Evol. *15*, e72027. https://doi.org/10.1002/ece3.72027.
33. Augustijnen, H., Patsiou, T., and Lucek, K. (2022). Secondary contact rather than coexistence—Erebia butterflies in the Alps. Evolution *76*, 2669–2686. https://doi.org/10.1111/evo.14615.
34. Kozak, K.M., Wahlberg, N., Neild, A.F.E., Dasmahapatra, K.K., Mallet, J., and Jiggins, C.D. (2015). Multilocus species trees show the recent adaptive radiation of the mimetic Heliconius butterflies. Syst. Biol. *64*, 505–524. https://doi.org/10.1093/sysbio/syv007.
35. Schmitt, T., and Hewitt, G.M. (2004). The genetic pattern of population threat and loss: a case study of butterflies. Mol. Ecol. *13*, 21–31. https://doi.org/10.1046/j.1365-294X.2004.02020.x.
36. Ehlers, S., and Schulz, S. (2023). The scent chemistry of butterflies. Nat. Prod. Rep. *40*, 794–818. https://doi.org/10.1039/D2NP00067A.
37. Liénard, M.A., Wang, H.-L., Lassance, J.-M., and Löfstedt, C. (2014). Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. Nat. Commun. *5*, 3957. https://doi.org/10.1038/ncomms4957.
38. Ômura, H., Noguchi, T., and Ohta, S. (2020). The male swallowtail butterfly, *Papilio polytes*, uses cuticular hydrocarbons for mate discrimination. Anim. Behav. *170*, 133–145. https://doi.org/10.1016/j.anbehav.2020.10.016.
39. Araújo, M.B., and Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. Glob. Ecol. Biogeogr. *16*, 743–753. https://doi.org/10.1111/j.1466-8238.2007.00359.x.
40. Catalan Butterfly Monitoring Scheme (2026). Clouded Apollo. https://www.catalanbms.org/en/especies/pappmn/.
41. Bolotov, I.N., Gofarov, M.Y., Rykov, A.M., Frolov, A.A., and Kogut, Y.E. (2013). Northern boundary of the range of the clouded Apollo butterfly *Parnassius mnemosyne* (L.) (Papilionidae): climate influence or degradation of larval host plants? Nota Lepidopterol. *36*, 19–33.
42. Franzén, M., and Johansson, V. (2025). Recent butterfly extinctions in Sweden reveal the inadequacy of site-based protection and the need for landscape-scale management. Conserv. Sci. Pract. *7*, e70188. https://doi.org/10.1111/csp2.70188.
43. Westin, A., Lennartsson, T., and Björklund, J.-O. (2018). The historical ecology approach in species conservation—identifying suitable habitat management for the endangered clouded Apollo butterfly (*Parnassius mnemosyne*) in Sweden. AIMS Environ. Sci. *5*, 244–272. https://doi.org/10.3934/environsci.2018.4.244.
44. Condamine, F.L. (2018). Limited by the roof of the world: mountain radiations of Apollo swallowtails controlled by diversity-dependence processes. Biol. Lett. *14*, 20170622. https://doi.org/10.1098/rsbl.2017.0622.
45. Tian, X., Mo, S., Liang, D., Wang, H., and Zhang, P. (2023). Amplicon capture phylogenomics provides new insights into the phylogeny and evolution of alpine Parnassius butterflies (Lepidoptera: Papilionidae). Syst. Entomol. *48*, 571–584. https://doi.org/10.1111/syen.12591.
46. Zhao, Y., He, B., Tao, R., Su, C., Ma, J., Hao, J., and Yang, Q. (2022). Phylogeny and biogeographic history of Parnassius butterflies (Papilionidae: Parnassiinae) reveal their origin and deep diversification in West China. Insects *13*, 406. https://doi.org/10.3390/insects13050406.
47. Eaton, D.A.R., and Overcast, I. (2020). ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics *36*, 2592–2594. https://doi.org/10.1093/bioinformatics/btz966.
48. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
49. Chang, C.C., Chow, C.C., Tellier, L.C.A., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.
50. R Core Team (2021). R: a Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
51. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.
52. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959. https://doi.org/10.1093/genetics/155.2.945.
53. Earl, D.A., and vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website for visualizing STRUCTURE output. Conserv. Genet. Resour. *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7.
54. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274. https://doi.org/10.1093/molbev/msu300.

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

55. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

56. Rambaut, A., and Drummond, A.J. (2012). Molecular evolution, phylogenetics and epidemiology. FigTree v1.4.2. http://tree.bio.ed.ac.uk/software/figtree/.

57. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252. https://doi.org/10.1093/molbev/msr048.

58. Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite-Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. *21*, 584–595. https://doi.org/10.1111/1755-0998.13265.

59. Chamberlain, S.A., and Boettiger, C. (2017). R, Python, and Ruby clients for GBIF species occurrence data. Preprint at PeerJ. https://doi.org/10.7287/peerj.preprints.3304v1.

60. Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B., and Anderson, R.P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography *38*, 541–545. https://doi.org/10.1111/ecog.01132.

61. Brown, J.L., Hill, D.J., Dolan, A.M., Carnaval, A.C., and Haywood, A.M. (2018). PaleoClim: a tool for generating high-resolution paleoclimate surfaces for global land areas. Ecol. Evol. *8*, 6354–6369.

62. Thuiller, W., Georges, D., Engler, R., and Breiner, F. (2023). biomod2: ensemble platform for species distribution modeling. Ecography *46*, e06014.

63. Gutenkunst, R.N., Hernández, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLOS Genet. *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695.

64. Portik, D.M., Leaché, A.D., Rivera, D., Barej, M.F., Burger, M., Hirschfeld, M., Rödel, M.O., Blackburn, D.C., and Fujita, M.K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. Mol. Ecol. *26*, 5245–5263. https://doi.org/10.1111/mec.14266.

65. Adams, D.C., and Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol. Evol. *4*, 393–399. https://doi.org/10.1111/2041-210X.12035.

66. Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. *13*, e1005752. https://doi.org/10.1371/journal.pcbi.1005752.

67. Brooks, M.E., Kristensen, K., Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., and Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. *9*, 378–400. https://doi.org/10.32614/RJ-2017-066.

68. Adams, D.C., Rohlf, F.J., and Slice, D.E. (2013). A field comes of age: Geometric morphometrics in the 21st century. Hystrix *24*, 7–14.

69. Mitteroecker, P., and Schaefer, K. (2022). Thirty years of geometric morphometrics: achievements, challenges, and the ongoing quest for biological meaningfulness. Am. J. Biol. Anthropol. *178*, 181–210. https://doi.org/10.1002/ajpa.24531.

70. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and parsing STRUCTURE results. Mol. Ecol. Resour. *15*, 1179–1191. https://doi.org/10.1111/1755-0998.12387.

71. Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., and Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. Sci. Data *4*, 170122. https://doi.org/10.1038/sdata.2017.122.

72. Keightley, P.D., Pinharanda, A., Ness, R.W., Simpson, F., Dasmahapatra, K.K., Mallet, J., Davey, J.W., and Jiggins, C.D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. Mol. Biol. Evol. *32*, 239–243. https://doi.org/10.1093/molbev/msu302.

73. Rohlf, F.J., and Slice, D.E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. *39*, 40–59. https://doi.org/10.2307/2992207.

74. Schmitt, T., Habel, J.C., Zimmermann, M., and Müller, P. (2006). Genetic differentiation of the marbled white butterfly, *Melanargia galathea*, accounts for glacial distribution patterns and postglacial range expansion in southeastern Europe. Mol. Ecol. *15*, 1889–1901. https://doi.org/10.1111/j.1365-294X.2006.02900.x.

75. Bookstein, F.L. (1991). Morphometric Tools for Landmark Data: Geometry and Biology (Cambridge University Press).

76. Collyer, M.L., and Adams, D.C. (2018). RRPP: an R package for fitting linear models to high-dimensional data using residual randomization. Methods Ecol. Evol. *9*, 1772–1779. https://doi.org/10.1111/2041-210X.13029.

77. Adams, D.C., and Collyer, M.L. (2024). Extending phylogenetic regression models for comparing within-species patterns across the tree of life. Methods Ecol. Evol. *15*, 2234–2246. https://doi.org/10.1111/2041-210X.14438.

78. Collyer, M.L., and Adams, D.C. (2024). Interrogating random and systematic measurement error in morphometric data. Evol. Biol. *51*, 179–207. https://doi.org/10.1007/s11692-024-09627-6.

79. Baken, E.K., Collyer, M.L., Kaliontzopoulou, A., and Adams, D.C. (2021). geomorph v4.0 and gmShiny: enhanced analytics and a new graphical interface for a comprehensive morphometric experience. Methods Ecol. Evol. *12*, 2355–2363. https://doi.org/10.1111/2041-210X.13723.

80. Jiggins, F.M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite-mediated speciation. Evolution *57*, 125–137.

81. Werren, J.H., Baldo, L., and Clark, M.E. (2008). *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. *6*, 741–751. https://doi.org/10.1038/nrmicro1969.

82. Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729. https://doi.org/10.1101/gr.210641.116.

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
42 *Parnassius mnemosyne* samples	Institut de Biologia Evolutiva (IBE)	Table S1
70 *Parnassius turatii samples*	Institut de Biologia Evolutiva (IBE)	Table S1
2 *Parnassius ariadne* samples	Institut de Biologia Evolutiva (IBE)	Table S1
Deposited data		
Parnassius mnemosyne and *Parnassius turatii* ddRADseq raw reads	This study	NCBI bioproject Nº: PRJNA1474983
Wing scale images of *Parnassius mnemosyne* for geometric morphometric analysis and the R script used for the analysis	This study	https://doi.org/10.5281/zenodo.20542681
CHC data and R script for CHC analysis	This study	https://doi.org/10.5281/zenodo.20544798
Software and algorithms		
IPYRAD v0.7.23	Eaton & Overcast[47]	https://ipyrad.readthedocs.io/en/master/
VCFtools v0.1.17	Danecek et al.[48]	https://vcftools.sourceforge.net/
PLINK v1.90b6.24	Chang et al.[49]	https://www.cog-genomics.org/plink/
R	R core team[50]	https://www.r-project.org/
Tidyverse (R package)	Wickham et al.[51]	https://github.com/tidyverse
STRUCTURE v2.3.4	Pritchard et al.[52]	https://web.stanford.edu/group/pritchardlab/structure.html
CLUMPAK v1.1	Kopelman et al.[52]	https://clumpak.evolseq.net/
STRUCTURE HARVESTER v0.6.94	Earl et al.[53]	https://github.com/StuntsPT/Structure_threader/blob/master/structure_threader/evanno/structureHarvester.py
IQTREE v2.1.2	Nguyen et al.[54]	https://github.com/iqtree/iqtree2
ModelFinder	Kalyaanamoorthy et al.[55]	N/A
FIGTREE v1.4.2	Rambaut et al.[56]	http://tree.bio.ed.ac.uk/software/figtree/
D-statistics	Durand et al.[57]	N/A
DSUITE	Malinsky et al.[58]	https://github.com/millanek/Dsuite
Rgbif R package v3.7.2	Chamberlain et al.[59]	10.32614/CRAN.package.rgbif
spThin R package	Aiello-Lammens et al.[60]	10.32614/CRAN.package.spThin
Paleoclim database	Brown et al.[61]	https://www.paleoclim.org/
biomod2 R package v3.5.1	Thuiller et al.[62]	10.32614/CRAN.package.biomod2
Dadi	Gutenkunst et al.[63]	https://dadi.readthedocs.io/en/latest/
dadi_pipeline v3.1.5	Portik et al.[64]	https://github.com/dportik/dadi_pipeline
easySFS	Gutenkunst et al.[62]	N/A
Geomorph R package	Adams et al.[65]; Baken et al.[66]	10.32614/CRAN.package.geomorph
mixOmics R package	Rohart et al.[66]	https://doi.org/10.5281/zenodo.1234
glmmTMB R package	Brooks et al.[67]	10.32614/CRAN.package.glmmTMB
Centrifuge v1.0.4	Kim et al.[67]	https://github.com/DaehwanKimLab/centrifuge

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

CellPress
OPEN ACCESS

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Sampling

P. mnemosyne is a butterfly species with a broad Palearctic distribution, spanning from Kazakhstan in the east to the Pyrenees in the west. Its range extends northward to Russia and southward to Iran, Turkey, the Balkans, and Italy. To capture the genetic diversity across this wide distribution, we sequenced 66 of the 122 available specimens from key regions, including the Balkans, Italy, Switzerland, Spain, Sweden, Ukraine, and Russia, among others. Of these, 29 specimens originated from west of the contact zone described by Bolotov et al.,[13] representing populations potentially corresponding to *P. turatii*, while the remaining specimens were sampled from east of the contact zone, likely representing *P. mnemosyne* sensu stricto. Additionally, two specimens of *Parnassius ariadne* were used as outgroup taxa. Butterfly bodies were stored in 99% ethanol at −20°C, and wings were kept separately as vouchers. While more specimens were initially used, some were excluded due to insufficient sequencing coverage in later steps (See Table S1).

METHOD DETAILS

ddRADseq library preparation

To obtain genome-wide representation, we applied a double-digest restriction-associated DNA (ddRAD) sequencing protocol. Genomic DNA (gDNA) was extracted from half of the thorax using the DNeasy Blood & Tissue Kit (Qiagen), and its quantity was assessed with the PicoGreen kit (Molecular Probes). Whole-genome amplification was performed using the REPLI-g Mini Kit (Qiagen) to enhance gDNA yield, followed by a second quantification with the PicoGreen kit. Each sample was digested using a combination of restriction enzymes: 500 ng of DNA was treated with 1 μL PstI, 2 μL MseI, and 5 μL of CutSmart Buffer (New England Biolabs), with water added to reach a final volume of 50 μL. Digestion was conducted at 37°C for 2 hours, after which samples were frozen to deactivate the enzymes. DNA purification was carried out using AMPure XP magnetic beads (Agencourt) on a Biomek automated liquid handler (Beckman Coulter), with a final elution volume of 40 μL. DNA concentrations were reassessed with PicoGreen, and these values were used for sample pooling. Adapter ligation was performed to enable individual sample identification. Each reaction included 5 μL T4 DNA Ligase Buffer (New England Biolabs), 1 μL T4 DNA Ligase (New England Biolabs), 0.6 μL rATP (Promega), 5 μL P1 adapter (50 nM), 5 μL P2 adapter (50 nM), and 2.4 μL water. The P1 adapter contained 45 unique Illumina primer sequences, 5 bp barcodes, and a TGCA overhang complementary to the PstI sticky end. The P2 adapter included Illumina primer sequences with AT overhangs to match the MseI cut site and incorporated a 'divergent-Y' structure to prevent amplification of fragments flanked by MseI cut sites. The ligation reaction was incubated at 22°C for 1 hour, followed by enzyme deactivation at 65°C for 20 minutes. For library preparation, 200 ng of each sample was pooled into three groups, each with a final volume of approximately 450 μL. The pooled samples underwent purification with AMPure XP magnetic beads, followed by size selection at 300 bp using the BluePippin system (Sage Science). Subsequently, PCR amplification was conducted using primers RAD1.F (5′ -AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG-3′) and RAD2.R (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC-3′). The reaction mixture (60 μL total volume) consisted of 9 μL water, 30 μL Phusion High-Fidelity PCR Master Mix (Finnzymes), 3 μL of each primer (10 mM), and 15 μL DNA. The PCR conditions included an initial denaturation at 98°C for 30 seconds, followed by 16 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 40 seconds, with a final extension at 72°C for 5 minutes. The amplified products were purified using AMPure XP magnetic beads, and DNA concentrations were re-evaluated with PicoGreen. Final size distribution and concentration of the pooled libraries were assessed using a Bioanalyzer (Agilent Technologies). Equimolar amounts of the libraries were pooled and sequenced on an Illumina HiSeqXten PE 150 bp platform at Macrogen The resulting demultiplexed FASTQ data were deposited in the NCBI Sequence Read Archive (SRA) associated with BioProject PRJNA1474983.

Wing scale imaging

To characterize the shape of wing scales we used geometric morphometrics.[68,69] Wing photographs were taken under a standardized setup using an Olympus SZX16 stereomicroscope equipped with an SDF PLAPO 1XPF objective. Images were acquired at a constant magnification of ×11.5 and under consistent illumination conditions provided by a Euromex stereomicroscope light source. Wing scales were flattened by placing them between two coverslips. An ocular micrometer (0–100 μm scale) was used for calibration and scale consistency across specimens.

CHC Extraction and GC-MS Analysis

For each specimen, the basal half of the hindwing was cut using clean scissors and immersed in 200 μL of pentane for 10 min at room temperature. The solvent was then evaporated to dryness under a gentle stream of nitrogen. The extract was resuspended in pentane, and 2 μL of this solution was injected into the Gas Chromatography-Mass Spectrometry (GC-MS).

Chemical analyses were performed using an Agilent 7820A gas chromatograph coupled to an HP 5977B mass selective detector (Agilent Technologies, Palo Alto, CA, USA). The GC was equipped with a non-polar capillary column (J&W DB-5ms fused silica, 30 m × 0.25 mm x 0.10 μm, Agilent Technologies). Injections were performed in splitless mode (1 min purge valve off). The injector port was set at 280°C, and the carrier gas was helium (at 15 psi head pressure). The oven temperature program was as follows: 70°C

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

held for 3 min, increased by 15°C/min to 150°C (held for 3 min), and then increased by 5°C/min to a final temperature of 310°C (held for 4.667 min). The mass spectrometer was operated in electron impact (EI) ionization mode at 70 eV. The transfer line, ion source, and quadrupole temperatures were set at 300°C, 230°C, and 150°C, respectively. Mass spectra were acquired in full scan mode in the range of 50–550 m/z. Data acquisition and analysis were performed using the Agilent Masshunter Workstation software (version B.07.00).

Compound identification was achieved by comparing the mass spectra of the eluted peaks with those present in the NIST library and by analyzing their fragmentation patterns. To confirm the identity of linear alkanes and calculate retention indices, a standard mixture of linear alkanes was injected under the same analytical conditions. We focused our analysis specifically on cuticular hydrocarbons; therefore, non-CHC compounds and contaminants were excluded from the dataset. Peak areas were integrated to estimate relative compound abundance. To ensure dataset consistency, peaks that co-eluted in some samples but appeared separated in others were summed and treated as a single variable.

An initial set of 43 CHCs was identified (See Table S6). Since the integration parameters were set to exclude negligible signals (implicitly filtering out peaks with <1% relative abundance), no further post-hoc abundance filter was applied. However, to ensure robustness, only compounds found in at least 25% of the samples were retained. This final filtering step resulted in a dataset of 30 CHCs used for subsequent statistical analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

RADseq data processing

To process the sequencing data, we employed IPYRAD v0.7.23[46] for initial filtering, SNP detection, and alignment. The genome of *P. mnemosyne* (*Parnassius_mnemosyne*_n_2023_11; GCA_963668995.1) was used as a reference, obtained from the ERGA project.[17] Key parameters were customized to suit the study's requirements: the assembly mode was configured to "reference-based," the data type was set to "pairddrad," and the restriction enzyme overhangs were specified as TGCAG and TAA. Additional adjustments included setting a minimum read length of 70 base pairs after trimming, requiring loci to be present in at least four samples, and limiting both the number of SNPs and indels per locus to 20 and 8, respectively. The resulting dataset consisted of 91,036 loci spanning a total of 19,837,673 base pairs, which included 1,514,746 high-quality SNPs.

Population genomics analysis

We assessed population genetic structure using Principal Component Analysis (PCA) and Bayesian clustering. The PCA was performed after filtering the dataset to exclude samples of *P. ariadne*. Additional filtering steps were conducted with VCFtools v0.1.17,[48] where we removed indels, excluded SNPs with more than 50% missing data, and retained variants with a minimum allele frequency of 0.008. To focus on unlinked SNPs, we applied variant pruning using PLINK v1.90b6.24,[49] with parameters set to a window size of 50 SNPs, a step size of 10 SNPs, and an r^2 threshold of 0.1. These steps yielded a final dataset of 15,041 SNPs. PCA computation was performed in PLINK and visualized in R[50] using the tidyverse package.[51]

For Bayesian clustering, we analyzed genetic structure with STRUCTURE v2.3.4.[52] We tested values of K ranging from 1 to 6, using a burn-in of 75,000 iterations followed by 200,000 Markov Chain Monte Carlo (MCMC) replicates. Each K was run ten times, and the outputs were averaged using CLUMPAK v1.1.[70] The optimal K was determined with the Evanno method in STRUCTURE HARVESTER v0.6.94.[53]

To infer phylogenetic relationships among the 63 individuals in the dataset, we constructed a maximum likelihood phylogeny using ^{IQTREE} v2.1.2.[54] Phylogenetic tree reconstruction was performed using a less stringently filtered dataset containing approximately 484,000 phylogenetically informative SNPs. The analysis employed the TVM+F+R2 substitution model, which was selected automatically by ModelFinder,[55] and included 1,000 bootstrap replicates. The resulting phylogeny was visualized, and bootstrap support was assessed with FIGTREE v1.4.2.[56]

Gene flow analysis

To investigate patterns of differential introgression between the populations belonging to taxa *mnemosyne* and *turatii*, we conducted an ABBA-BABA analysis (also referred to as D-statistics[57]). The analysis was performed using a variant calling file containing 384,718 SNPs from a dataset that had been filtered to exclude contaminants. Calculations were carried out using the DTRIOS module within the DSUITE software package.[58] DTRIOS employs a standard block-jackknife procedure to evaluate whether the D-statistic deviates significantly from zero, which represents the null hypothesis of no introgression. The ordering of P1 and P2 ensures that nABBA $\geq$ nBABA, meaning that a significant result would identify the species assigned as P2 as the one potentially introgressed.

We analyzed a dataset of 384,718 variants using Dsuite Dtrios. The analysis included three sets (excluding the outgroup), and D and f_4f_4-ratio statistics were calculated for four trio combinations. Two of these trios focused on populations in the contact zone, while the other two targeted potential introgression events suggested by the structure analysis and PCA.

Ecological Niche Modeling

We compiled occurrence data for *P. mnemosyne* sensu lato from the Global Biodiversity Information Facility (GBIF) covering the years 1979 to 2013. The dataset contained 6,758 records, of which 3,218 were assigned to the taxon *mnemosyne* and 3,540 to the taxon *turatii* based on the lineage distributions described by Bolotov et al.[13] and analyzed separately. Data were downloaded

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

using the occ_download function from the rgbif R package (v3.7.2[59]), specifying a geographic region bounded by longitudes −11.64551° to 48.20801° and latitudes 29.26758° to 72.1582°. We excluded records with geospatial issues and restricted thbueno e query to this time frame. To reduce spatial sampling bias, we applied a thinning procedure using the spThin R package, maintaining a minimum distance of 10 km between occurrence points.

Environmental predictor variables were selected to capture key climatic factors shaping species distributions. We used 19 bioclimatic variables at ∼10 km resolution (5 arc-minutes) from the Paleoclim database[61] for the present (1979–2013), as well as paleoclimatic data from the Last Glacial Maximum (21,000 years ago[71]). To address collinearity,[40] we calculated Pearson correlation coefficients and retained only variables with correlations below 0.7. The final predictor set included mean diurnal range (bio2), minimum temperature of the coldest month (bio6), mean temperature of the wettest quarter (bio8), precipitation of the coldest quarter (bio19), and precipitation of the warmest quarter (bio18), prepared as raster layers for modeling.

Species distribution modeling was performed using the biomod2 R package (v3.5.1[62]) to estimate habitat suitability under present and historical climates. We complemented presence data with pseudo-absence points by randomly selecting five independent sets of 1,000 points across the study area to reflect the environmental background. Six modeling algorithms were used: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Multiple Adaptive Regression Splines (MARS), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), and Random Forests (RF). We trained models on 80% of the occurrence data and assessed performance using the True Skill Statistic (TSS), with ten cross-validation replicates to ensure robust predictions. Cross-validation indicated that all distribution models performed well, with mean validation TSS values of 0.882 for GLM, 0.913 for GAM, 0.887 for MARS, 0.938 for CTA, 0.954 for RF, and 0.893 for FDA.

For each time period, final habitat suitability maps were generated by averaging models with TSS > 0.7 to produce consensus projections. Importantly, all analyses were conducted separately for *P. mnemosyne* and *P. turatii* to explicitly compare their ecological niches and potential range differences.

Demographic Modeling Inference

We used the dadi software[63] and dadi_pipeline v3.1.5[64] to reconstruct demographic history and gene flow patterns among *Parnassius* populations based on the joint site frequency spectra (JSFS). The JSFS was built using the easySFS[63] tool from a filtered ddRAD-seq. To ensure data quality, we retained only unlinked SNPs and excluded loci with over 50% missing data, minimizing biases from linkage disequilibrium and low-quality genotypes.

We tested four demographic models: (1) split_uni_mig_adjacent_var2 (split with unidirectional migration between adjacent populations), (2) split_sym_mig_adjacent_var2 (split with symmetric migration), (3) admix_origin_no_mig (admixture origin without ongoing migration), and (4) split_nomig (pure isolation with no migration). Each model was evaluated twice, switching the population ordering in the input to test whether the Pyrenean population is more closely related to *P. turatii* or *P. mnemosyne*, directly addressing uncertainties in its phylogenetic placement. Model implementation and optimization were carried out using dadi_pipeline with four iterative rounds of parameter fitting. Replicates achieving the highest log-likelihoods were used to initialize subsequent rounds. Optimization relied on the Nelder–Mead algorithm (optimize_log_fmin), following dadi_pipeline's defaults: 10, 20, 30, or 40 replicates per round, with up to 3, 5, 10, or 15 iterations per replicate. Model selection was based on the Akaike Information Criterion (AIC), selecting the model with the lowest AIC as the best fit.

To convert estimated parameters—ν (relative population size), m (migration rate), and t (divergence time)—into biological units, we used the θ (theta) estimate from dadi. The ancestral effective population size (Nref) was calculated as $Nref = \theta / (4L\mu)$, where L is the number of analyzed sites, determined as the proportion of retained SNPs (31,255) relative to the total called SNPs (1,514,746 across 19,837,673 base pairs), and μ is the per-site per-generation mutation rate. We used a mutation rate of 2.9×10^{-9} per base pair per generation, based on *Heliconius melpomene*,[72] widely used in Lepidoptera. We assumed a generation time of 1 year (one generation per year). Final estimates included $N = \nu \times Nref$, migration rates per generation as $M = m / (2Nref)$, and divergence times in years as $T = 2t \times Nref \times g$, where g is generation time. These conversions provided biologically meaningful estimates of population sizes, migration rates, and divergence times, offering detailed insight into the evolutionary and demographic processes shaping the *P. mnemosyne* complex.

Wing scale morphometrics

We selected five scales from the wings of each of 42 individuals and digitized the locations of four landmarks demarcating the corners of each scale to represent its shape (Figure 4B). We then performed a Generalized Procrustes Analysis[73] to eliminate nonshape variation and align the 210 scales to a common coordinate system. The aligned Procrustes coordinates were then used in all statistical analyses.[74] Specifically, we performed a principal components analysis (PCA) to visualize overall patterns of scale shape variation. Thin-plate spline deformation grids[75] were then used to generate graphical representations of scale shapes for the means of each group. We then compared scale shapes across groups, using a permutation-based hierarchical multivariate analysis of variance. Residual randomization permutation procedures (RRPP[76]) were performed within individuals to account for the nonindependence among scales obtained from the same individual (for related procedures see Adams and Collyer[77]; Collyer and Adams[78]). Pairwise comparisons among groups were also conducted using the same permutation schedule.[76] Additional analyses comparing species distinctness based on scale shape are found in the Supplementary (See Table S4). All analyses were conducted in R, using the geomorph[65,79] and RRPP[76] packages.

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

Chemical discrimination and data preprocessing

Cuticular hydrocarbon (CHC) peak areas were used as compositional chemical profiles. To account for the compositional nature of the data, zero values were replaced using a count zero multiplicative (CZM) approach implemented in the cmultRepl function of the zCompositions package, and data were subsequently transformed into Aitchison geometry using the centred log-ratio (CLR) transformation (using the clr function of the compositions R package). Analyses were restricted to male specimens to avoid sex-related chemical variation.

For descriptive and exploratory purposes, specimens were assigned to a combination of species and geographic regions. In particular, we identified two contact populations—*P. turatii* from the Alps and *P. mnemosyne* from the same area—, defined as the populations within 100km from the known contact point (46.0 lat; 12.0 long). Non-contact populations for each species were the Pyrenees and Western Alps–Italian Peninsula-Sicily populations for *P. turatii*; and North-Eastern Europe populations for *P. mnemosyne*.

To assess whether a combination of chemical compounds could discriminate between *P. mnemosyne* and *P. turatii*, we performed sparse Partial Least Squares Discriminant Analysis (sPLS-DA) using the splsda function of the mixOmics R package.[66] Unlike unsupervised ordinations, sPLS-DA explicitly maximizes covariance between chemical profiles (X) and species identity (Y), and is therefore appropriate to test the a priori discriminability of species based on CHC composition. We fitted a two-component sPLS-DA model, retaining a maximum of ten compounds per component which assured to avoid overfitting since more than 10 specimens per variable in each component were included. SPLSDA allowed both dimensionality reduction and variable selection. Model performance on specimens from different species and regions was evaluated using leave-one-out cross-validation (LOO), implemented via the perf() function in mixOmics with centroid-based distances. For each specimen, posterior class membership probabilities were extracted and the predicted class was assigned as the class with the highest probability. Classification accuracy was calculated as the percentage of correctly classified specimens, both overall and separately for each geographic group. This allowed us to assess whether classification performance differed between contact and non-contact populations.

To obtain an unsupervised low-dimensional representation of chemical variation, we performed a principal component analysis (PCA) on the CLR-transformed data (prcomp, centred but not scaled). PCA was used to obtain an orthogonal representation of chemical variation, thereby reducing redundancy among correlated compounds before calculating interspecific distances. Scores on the first two principal components (PC1–PC2; or the set of PCs specified a priori) were used as a Euclidean embedding of chemical profiles. Pairwise chemical distances were calculated as Euclidean distances between individuals in this PCA space.

For each individual, geographic distance to the contact point (46.0 lat; 12.0 long) was computed using great-circle distances (earth.dist from the fossil package). This yielded an individual-level predictor variable (dist_contact, in km).

As an exploratory analysis based on complete pairwise distance matrices, we performed a Multiple Regression on distance Matrices (MRM; ecodist package). Chemical distances were calculated as Euclidean distances between individuals in PCA space using the first eight principal components, which together explained more than 70% of the total variance. Geographic distances between individuals were calculated as great-circle distances from sampling coordinates. In addition, we constructed a binary pair-type matrix distinguishing conspecific (0) from heterospecific (1) comparisons.

The chemical distance matrix, geographic distance matrix, and pair-type matrix were converted into aligned vectors containing all pairwise comparisons. We then fitted an MRM model with chemical distance as the response variable and geographic distance, pair type, and their interaction as predictors:

$$\text{Chemical distance} \sim \text{Geographic distance} + \text{Pair type} + \text{Geographic distance} \times \text{Pair type}$$

Statistical significance was assessed by permutation (9,999 permutations). This analysis was intended as a descriptive evaluation of whether the relationship between geographic and chemical divergence differed between conspecific and heterospecific pairs across the entire dataset. In particular, the interaction term tested whether geographic distance had different effects on chemical divergence within and between species.

To directly test whether interspecific chemical divergence changes as a function of distance from the contact zone, we performed a second analysis restricted to comparisons between species. Specifically, we extracted the rectangular block of the pairwise chemical distance matrix corresponding to all comparisons between *P. mnemosyne* and *P. turatii* individuals (mnemosyne × turatii; n = 2,600 pairwise comparisons). For each interspecific comparison, we retained the chemical distance in PCA space (chemDist) together with the distance from the contact zone of the focal individual (contactDist). By focusing exclusively on heterospecific comparisons, this approach specifically evaluates whether chemical differentiation between species is strongest near the contact zone and declines with increasing distance from it.

Because each individual contributes to multiple pairwise comparisons, we represented the data in two complementary sets: *mnemosyne*-as-focal: each *mnemosyne* individual paired with all *turatii* individuals, with contactDist defined as that *mnemosyne* individual's distance to the contact zone; and *turatii*-as-focal: each *turatii* individual paired with all *mnemosyne* individuals, with contactDist defined as that *turatii* individual's distance to the contact zone. These two sets were combined into a single dataset with four key variables: interspecific chemical distance (chemDist), distance from the contact zone (contactDist), focal species: *mnemosyne* vs *turatii* (Species) and ID of the focal specimen (id_focal). As each observation corresponds to a chemical distance between *P. mnemosyne* and *P. turatii*, species identity does not represent an independent grouping factor but only specifies which individual provides the distance to the contact zone; accordingly, models focused on distance effects rather than species main effect.

Please cite this article in press as: Lancho-Silva et al., Second-chance speciation in the Clouded Apollo butterfly, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.06.052

Pairwise distances are not independent because each focal individual appears repeatedly across multiple pairwise comparisons. To account for this non-independence, we fitted a Gaussian mixed model using the glmmTMB R package[67] with a random intercept for the focal individual:

$$\text{Chemical distance} \sim \text{Contact distance} + \text{Contact distance} : \text{Species} + (1|\text{id_focal})$$

Models were fitted using REML. The main effect of Contact distance tests whether interspecific Chemical distances with geographic distance from the contact zone, the main effect of Species tests for overall differences between focal-species views, and the interaction tests whether the distance–divergence relationship differs between the two focal species.

To visualize fitted trends, we generated marginal predictions of Chemical distances along the observed range of Contact distances , separately for each focal species. Predicted values and 95% confidence intervals were plotted in a mirrored layout: the *P. turatii* trend was displayed on the negative x-axis (left), and the *P. mnemosyne* trend on the positive x-axis (right), with the contact zone centred at x = 0. This representation allows direct visual comparison of distance-dependent trends between species while keeping the contact point as a common reference.

Wolbachia infection

To assess the potential influence of *Wolbachia* on genetic differentiation, we screened for this endosymbiotic bacterium, which is known to cause reproductive alterations such as male-killing and cytoplasmic incompatibility.[80,81] Sequence data were analyzed using CENTRIFUGE v1.0.4,[82] a rapid and sensitive tool for bacterial sequence detection. Because some reads did not align to the *P. mnemosyne* reference genome, we performed a de novo assembly using IPYRAD v0.7.23[47] specifically to retain potential symbiont sequences that would otherwise be discarded. The assembled loci were then analyzed in CENTRIFUGE using the complete bacterial and viral genome index from the NCBI BLAST nucleotide database, enabling the identification and classification of *Wolbachia* sequences.