Convergence in sympatric swallowtail butterflies reveals ecological interactions as a key driver of worldwide trait diversification

Agathe Puissant, Ariane Chotard, Fabien Condamine, Violaine Llaurens

Convergence in sympatric swallowtail butterflies reveals ecological interactions as a key driver of worldwide trait diversification

Agathe Puissant[1], Ariane Chotard[1], Fabien L. Condamine[2], Violaine Llaurens[1]

Institut de Systématique, Evolution et Biodiversité (UMR 7205 CNRS/MNHN/SU/EPHE/UA), Muséum National d'Histoire Naturelle - CP50, 57 rue Cuvier, 75005, Paris, France

CNRS, Institut des Sciences de l'Évolution de Montpellier (Université de Montpellier), Place Eugène Bataillon, 34095, Montpellier, France

*Agathe Puissant

Email: agathe.carquainpuissant@gmail.com

Author Contributions: A.P. and V.L. designed research. A.P. developed the method and analyzed data. All authors performed research and wrote the paper.

Competing Interest Statement: The authors declare no competing interests.

Keywords: machine learning; Papilionidae; wing color pattern; macroevolution; mimicry.

Abstract

Ecological interactions can promote phenotypic diversification in sympatric species. While competition can enhance trait divergence, other ecological interactions may promote convergence in sympatric species. Within butterflies, evolutionary convergences in wing color patterns have been reported between distantly-related species, especially in females of palatable species, where mimetic color patterns are promoted by predator communities shared with defended species living in sympatry. Wing color patterns are also often involved in species recognition in butterflies, and divergence in this trait has been reported in closely-related species living in sympatry as a result of reproductive character displacement. Here, we investigate the effect of sympatry between species on the convergence vs. divergence of their wing color patterns in relation to phylogenetic distance, focusing on the iconic swallowtail butterflies (family Papilionidae). We developed a new unsupervised machine-learning-based method to estimate phenotypic distances among wing color patterns of 337 species, enabling us to finely quantify morphological diversity at the global scale among species, and allowing us to compute pairwise phenotypic distances between sympatric and allopatric species pairs. We found phenotypic convergence in sympatry, stronger among distantly related species, while divergence was weaker and restricted to closely-related males. The convergence was stronger among females than males, suggesting that differential selective pressures acting on the two sexes drove sexual dimorphism. Our results highlight the significant effect of ecological interactions driven by predation pressures on trait diversification in Papilionidae and provide evidence for the interaction between phylogenetic proximity and ecological interactions in sympatry, acting on macroevolutionary patterns of phenotypic diversification.

Significance Statement

Disentangling the effects of ecological and phylogenetic factors acting on trait evolution is challenging. Evolutionary convergences in color patterns are often documented for distantly-related species targeted by the same predators, promoting mimicry in prey species. Here, we investigated the global effect of sympatry on wing color pattern evolution

to assess the importance of ecological interactions on trait diversification. We developed a new machine learning tool to automatically quantify color pattern variations across the butterfly family Papilionidae. Living in sympatry clearly promoted convergence between distantly-related species, with a stronger effect on females, suggesting that differential selective pressures on the two sexes drove sexual dimorphism. We showcase the global signature of differential selective pressures on trait diversification across multiple species.

Introduction

Species interactions are an important driver of phenotypic diversification at both microevolutionary and macroevolutionary scales, promoting either divergence or convergence in sympatry. While local species communities are strongly shaped by contingent historical factors, they may also be influenced by both habitat filtering and ecological interactions (e.g. as documented in oak communities, 1), thus impacting the evolution of their respective traits. Indeed, antagonistic and mutualistic interactions fuel the evolution of suites of traits involved in adaptation to specialized ecological niches, therefore overcoming the effect of common ancestry. Traits involved in interspecific interactions have often been shown to diverge in coexisting species, due to their partitioning into different niches (2–6). Phenotypic convergence can also be observed in sympatric species at large scales because of shared selective pressures (e.g. in Ithominii butterflies, 7).

Importantly, the effect of species interactions on trait similarity within species communities depends on the degree of relatedness between sympatric species. For instance, interspecific competition for territory (4) or mates (8) can enhance phenotypic divergence between sympatric species, depending on their phylogenetic proximity. Indeed, closely-related species are more likely to occupy similar ecological niches and display a similar set of traits (9). Such resemblance may further increase interspecific sexual interactions and promote trait displacement in closely-related species (10, 11). Reproductive interferences may thus accelerate the evolutionary divergence of traits involved in species recognition between closely-related species living in sympatry (12, 13). Interspecific interactions such as mimicry, on the other hand, are more likely to filter species communities according to the similarity of their traits (7), so that species with ancestral similar color patterns are more likely to co-occur in sympatry. The advantage conferred by local resemblances within communities may also induce phenotypic convergence between phylogenetically-distant species (14). Here, we aim at testing for the relative effects of 1) convergence promoted by local adaptation vs. 2) divergence driven by competitive interactions on trait diversification

at the macroevolutionary scale. Because ecological interactions vary with the degree of phylogenetic proximity within species communities, we specifically test for the interactions between local ecological processes and phylogenetic relatedness to provide an assessment of the relative roles of historical and ecological factors as drivers of trait diversification at the global scale.

Within butterflies, convergence in wing color pattern can induce countervailing effects between ecological interactions and phylogenetic distances. In unpalatable species, the evolution of mimetic color patterns in sympatry is promoted by predator learning and avoidance of prey harboring a known warning signal. This predator behavior generates positive density-dependent selection that favors the convergence of wing color patterns in sympatry, referred to as Müllerian mimicry (15, 16). The evolution of mimetic color patterns is also commonly observed in palatable species, referred to as the Batesian mimicry, where displaying a coloration close to that of the defended prey confers a strong advantage (17). Both Müllerian and Batesian mimicries were observed in butterflies (e.g. in the chemically-defended *Heliconius* species (18) and in the palatable Papilio species (19)) A striking example is the strong resemblance that can be observed between individuals from the unpalatable *Papilio polytes* and the chemically-defended *Pachliopta aristolochiae* (20), while these clades diverged about 50 million years ago.

In addition, butterfly wing color pattern is involved in species recognition during mate choice, favoring divergence between closely-related species living in sympatry (21). This trade-off between natural and sexual selection is likely to have different consequences on the evolution of the color pattern in the two sexes. The slower flight of egg-loaded females, as well as their predictable behavior of laying eggs on specific host plants, may increase their predation risk and further promote Batesian mimicry in females (22, 23). Reproductive interference with other sympatric species may favor female preference for non-mimetic males, thereby promoting mimicry in females but not in males (24). Thus, the evolutionary trade-off between convergence on mimetic signals and divergence in mating cues may differ between sexes, driving the evolution of sexual dimorphism. What evolutionary force prevails on the diversification of butterfly color patterns? A global

pattern of sympatric divergence is expected, especially in males, if sexual selection linked to species recognition dominates, whereas recurrent convergence is predicted if natural selection imposed by predators plays a predominant role.

Here, we investigate the effect of sympatry on the macroevolutionary trend of color pattern diversification in swallowtail butterflies (Papilionidae). The Papilionidae family is recognized as a textbook example of the evolution of mimetic color patterns (19) and sexual dimorphism, its ecology and geographic distribution are well documented, and phylogenetic relationships between species are well established (25), making it an ideal model to test the effect of local ecological processes on the general trend of trait evolution. We used 1,358 photographed individuals from 337 species of this family distributed throughout the world. Specifically, we aim at testing the opposite effect of mimicry and reproductive interference, and its interaction with phylogenetic proximity between interacting species. Wing color patterns in butterflies are complex traits composed of different features, such as stripes, spots, rays of different shapes and colors, which may have evolved in concert (26). To assess convergence and divergence in these complex color patterns, we develop a novel machine learning-based method to quantify subtle variations in wing color patterns. This quantification of color pattern variation then allows estimating the effect of geographic overlap between species on the evolution of wing color patterns displayed by males and females. We specifically test whether sympatric species have significantly 1) more or 2) less similar color patterns than expected based on their common ancestry, allowing to detect evolutionary convergence and divergence, respectively. Significant convergence in sympatry may indeed be consistent with a role for predator-mediated selection promoting mimicry, while divergence may suggest an important effect of reproductive interference. Finally, because the relative effect of natural and sexual

selection may differ between sexes, we examine 3) the sex-specific effect of sympatry on color pattern evolution.

Results

1. Quantifying color pattern variation using an unsupervised method

We quantified color pattern similarity within and among Papilionidae species without relying on any pre-existing human classification by developing an unsupervised similarity learning algorithm, based on the SimCLR method (a Simple Framework for Contrastive Learning of Visual Representations, 27). This method allows an objective quantification of color pattern variation between distantly related species that may differ greatly in wing shape and venation (see Methods section for detailed comparisons with existing methods). Similarity learning relies on modifications of the original images: the algorithm places the modified versions from the same image close together in the representation space, while modified versions from different images stand at larger distances. This unsupervised method allows partial control over the features used during classification: by cropping and rotations images to produce the modified images, we forced the model to ignore wing shape variations between species, while color pattern variations are used as discriminative features.

We used 2,716 standardized photographs of the dorsal and ventral sides of Papilionidae butterflies and automatically separated the wings from the body. The model was then trained on the modified images of the four wings together standing on a neutral background. This training resulted in representation vectors of dimension 2,048 for each butterfly image, containing information about the features displayed in the image. We then used a Principal Component Analysis (PCA) to reduce the dimensionality to 20, while retaining approximately 80% of the variance. In certain species, several phenotypic forms

are described, so we carried out the examination at the form level. We plotted the mean phenotype for each form within species in the resulting morpho-space (Figure 1 a,b). As there is considerable sexual dimorphism in Papilionidae, we first performed independent analyses for males and females. As expected, interspecific distances were significantly much higher than intraspecific distances for both males and females, when studying either the dorsal or ventral side (Wilcoxon, p-value < 0.001 for males and p-value < 0.001 for females, Fig. S2), indicating that the model was able to successfully discriminate species based on phenotypic differences. Because the dorsal and ventral patterns were very similar for most species, we only show the analyses performed on the dorsal wing color patterns. While our method does not allow us to precisely identify the contribution of different wing features to the different PC axes, the first axis clearly discriminates between colored vs. white and black patterns. The gradient-based class activation mapping (Grad-CAM) for each picture then allows to pinpoint the pixels that were most used as discriminative features by the model (Figure 2). Elements of the pattern, such as stripes or color patches, show high activation, and were therefore used as discrimination criteria by the model. In most species displaying hindwing tails, the tails were not considered as discriminant features (e.g. *Battus philenor*, *Graphium weiskei*, *Pachliopta aristolochiae*, *Papilio caiguanabus*; Figure 2 c, d, g, h). However, in some species where the hindwing tails are particularly prominent (e.g. *Lamproptera curius*; Figure 2 m), the model did rely on the tails as a discriminant feature. Thus, our phenotypic quantification relied mostly on color pattern variation rather than wing shape, but still accounted for large, visually-discriminant, wing shape variations.

2. Phylogenetic signal

We then constructed a phylo-morphospace for each sex to describe the color pattern variation throughout the whole Papilionidae family (Figure 1 c,d). These morphospaces show the position of the average dorsal color pattern within each species, and a projection of the phylogenetic relationships among species using the principle of unscaled squared change parsimony, based on the most recent and complete swallowtail phylogeny (25). We

detected a significant phylogenetic signal on phenotypic variation in both males ($K = 0.39$, p-value < 0.001) and females ($K = 0.32$, p-value < 0.001). K values less than 1 for both sexes indicate that closely-related species were generally less phenotypically similar than expected under a Brownian motion model of phenotypic evolution. We fitted several models of multivariate trait evolution using the mvMorph package (28) and found that a lambda model, which accounts for the degree of phylogenetic signal, best fit our data (supplementary information 4).

Mimicry between species, and especially female-limited mimicry documented in Papilionidae, is likely to account for the limited phylogenetic signal in color pattern evolution. To test the effect of interspecific mimicry on the macroevolutionary pattern of

wing color pattern diversification, we investigated the effect of sympatry between species on color pattern evolution depending on the level of phylogenetic relatedness.

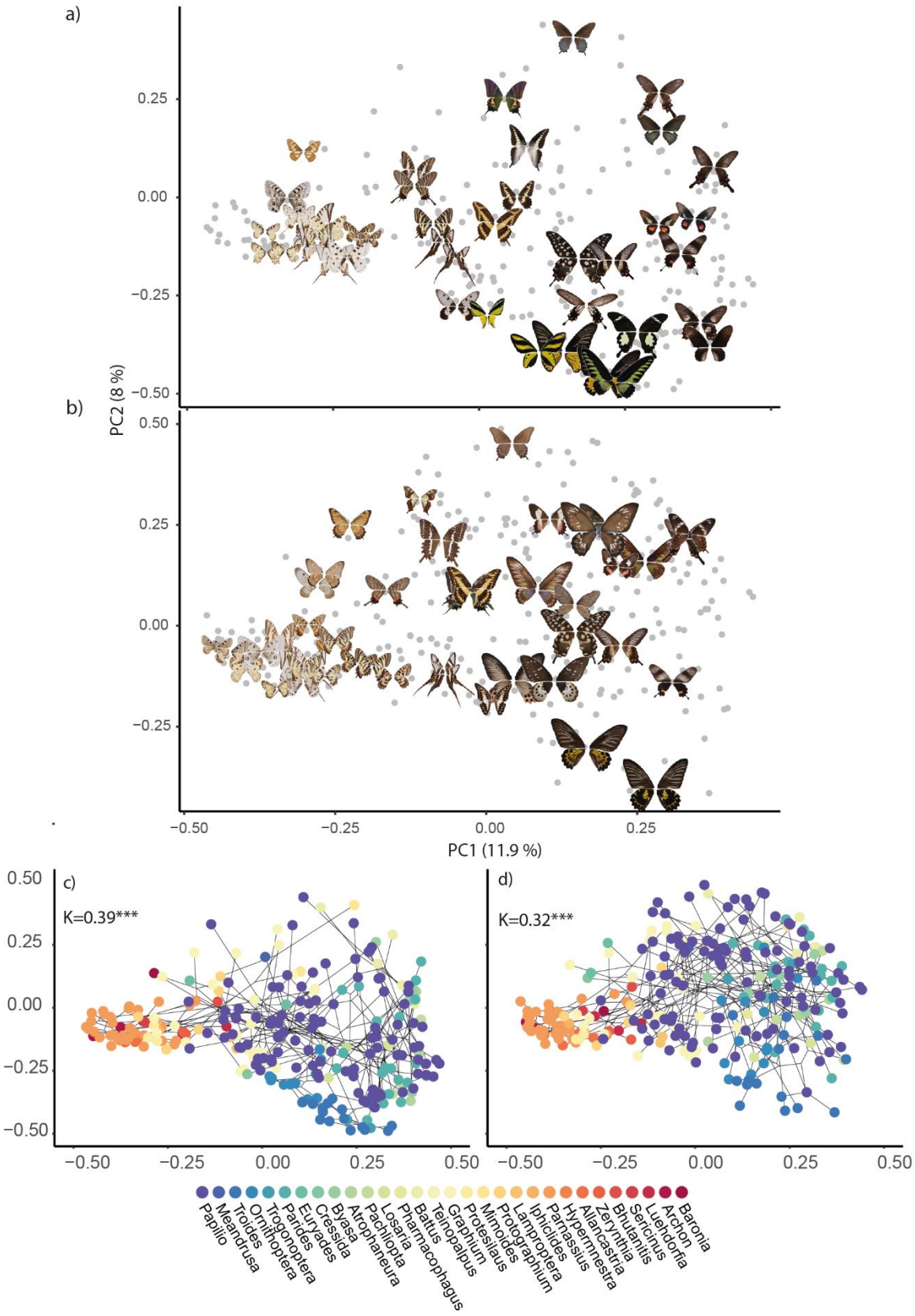
a)
b)
PC2 (8 %)
PC1 (11.9 %)
0.25
0.00
−0.25
−0.50
c)
K=0.39***
d)
K=0.32***
Papilio
Meandrusa
Troides
Ornithoptera
Trogonoptera
Parides
Euryades
Cressida
Byasa
Atrophaneura
Pachliopta
Losaria
Pharmacophagus
Battus
Teinopalpus
Graphium
Protesilaus
Mimoides
Protographium
Lamproptera
Iphiclides
Parnassius
Hypermnestra
Allancastria
Zerynthia
Bhutanitis
Sericinus
Luehdorfia
Archon
Baronia

Figure 1. Color pattern variations and phylogenetic relationships across Papilionidae butterflies. First row: Phenotypic variations captured by the unsupervised machine learning-based methods applied to our 2,716 pictures of Papilionidae. Independent PCA were carried out on (a) males and (b) females. Note that we used the mean phenotype by sex and by species and represented only the first two axes, explaining only 19.9% of the phenotypic variance. We display the actual picture of butterflies for randomly sampled species (with at least one species per genera) on the morphospace to observe how actual color pattern variation was separated by our method of phenotypic discrimination. Second row: Phylo-morpho space computed on the mean phenotype for each species in (c) males and (d) females. The phylo-morpho space is the projection of the morpho space coordinates on the first two axes of the morpho space principal components. Each colored dot represents the phenotype of a given species and the black lines show the projection of the

phylogenetic relationships among the species. The color code corresponds to the different genera, the color gradient corresponding to the location of the genus on the phylogeny.

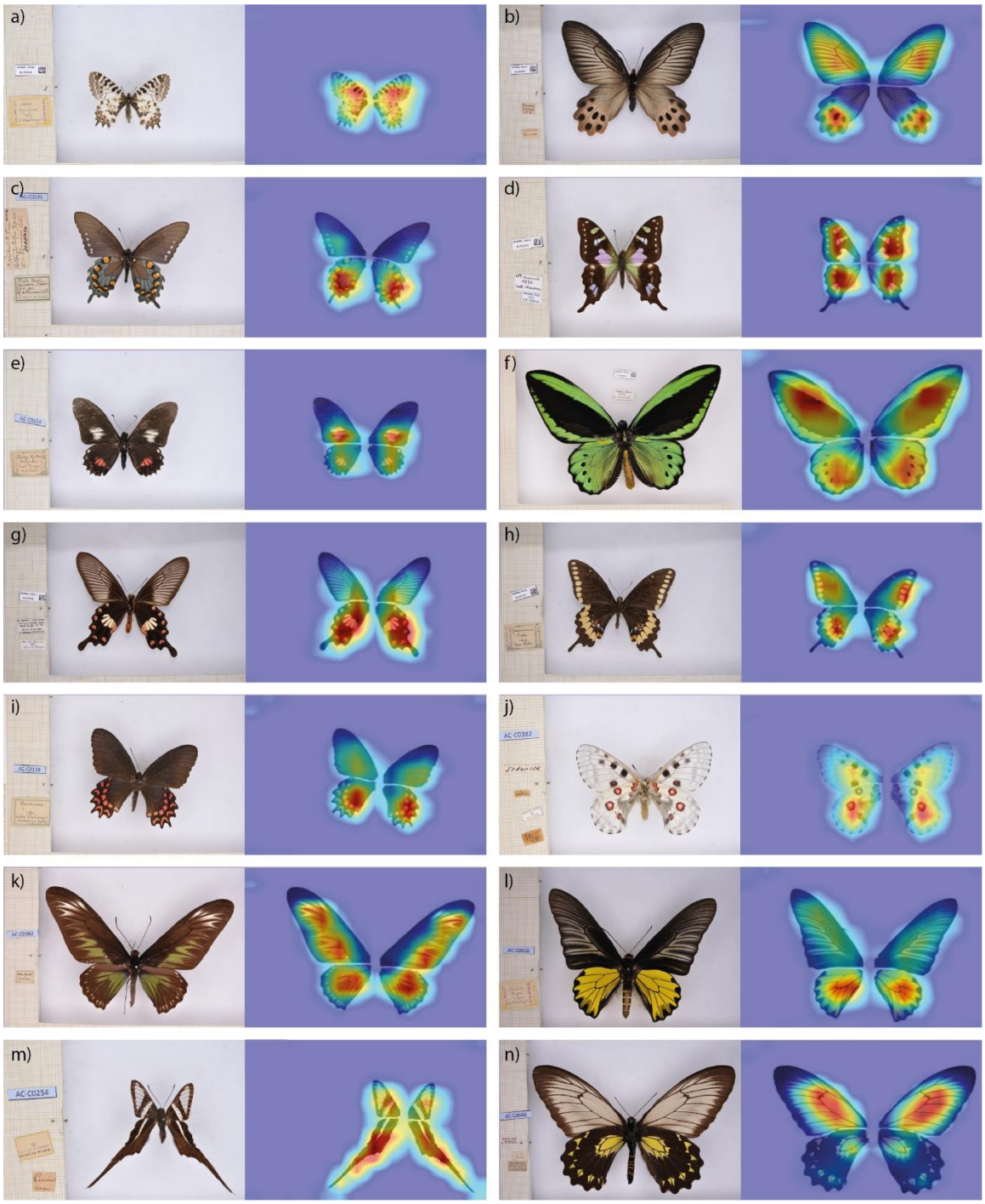

Figure 2. Discriminant features used to build the morpho-space obtained using Grad-CAM mapping. We selected one or two species for several genera to give examples that cover different parts of the phenotypic space. Redder color indicates pixels that weigh the most in the activation of the neural network. From left to right, up to down: *Allancastria cerisyi*, *Atrophaneura priapus*, *Battus philenor*, *Graphium weiskei*, *Mimoides euryleon*, *Ornithoptera priamus*, *Pachliopta aristolochiae*, *Papilio caiguabanus*, *Parides photinus*, *Parnassius nomion*, *Trogonoptera brookiana*, *Troides aeacus*, *Lamproptera curius*, *Troides andromache*.

3. Color pattern convergence is more frequent in sympatric species and increases with geographic overlap

To disentangle the complex interactions between historical and ecological factors acting on phenotypic evolution, we fitted phylogenetic linear mixed models (PLMM) with phenotypic distance as the response variable and percentage of geographic overlap between species, phylogenetic distance and phylogenetic distance squared and their interaction as the predictors. We found a significant negative effect of the degree of geographic overlap on phenotypic distance (coefficient -0.012 p-value<0.001 for males and coefficient -0.015 p-value<0.001 for females), meaning that as the degree of sympatry increases, phenotypic distance tends to decrease. We also found significant interactions between the geographic overlap and the phylogenetic distance. Full model results can be found in table S3 and S4 in the supplementary information. Figure 3 then shows the variation in the relationship between overlap and phenotypic distance for three ranges of different phylogenetic distances (12.1 Myr, 51.87 Myr and 91.63 Myr - sum of branch lengths). Overall the slope of phenotypic distance in function of overlap becomes more negative with increasing phylogenetic distance for both males and females. However, for males, the slope is positive

at small phylogenetic distances, suggesting phenotypic divergence in males from closely-related species living in sympatry.

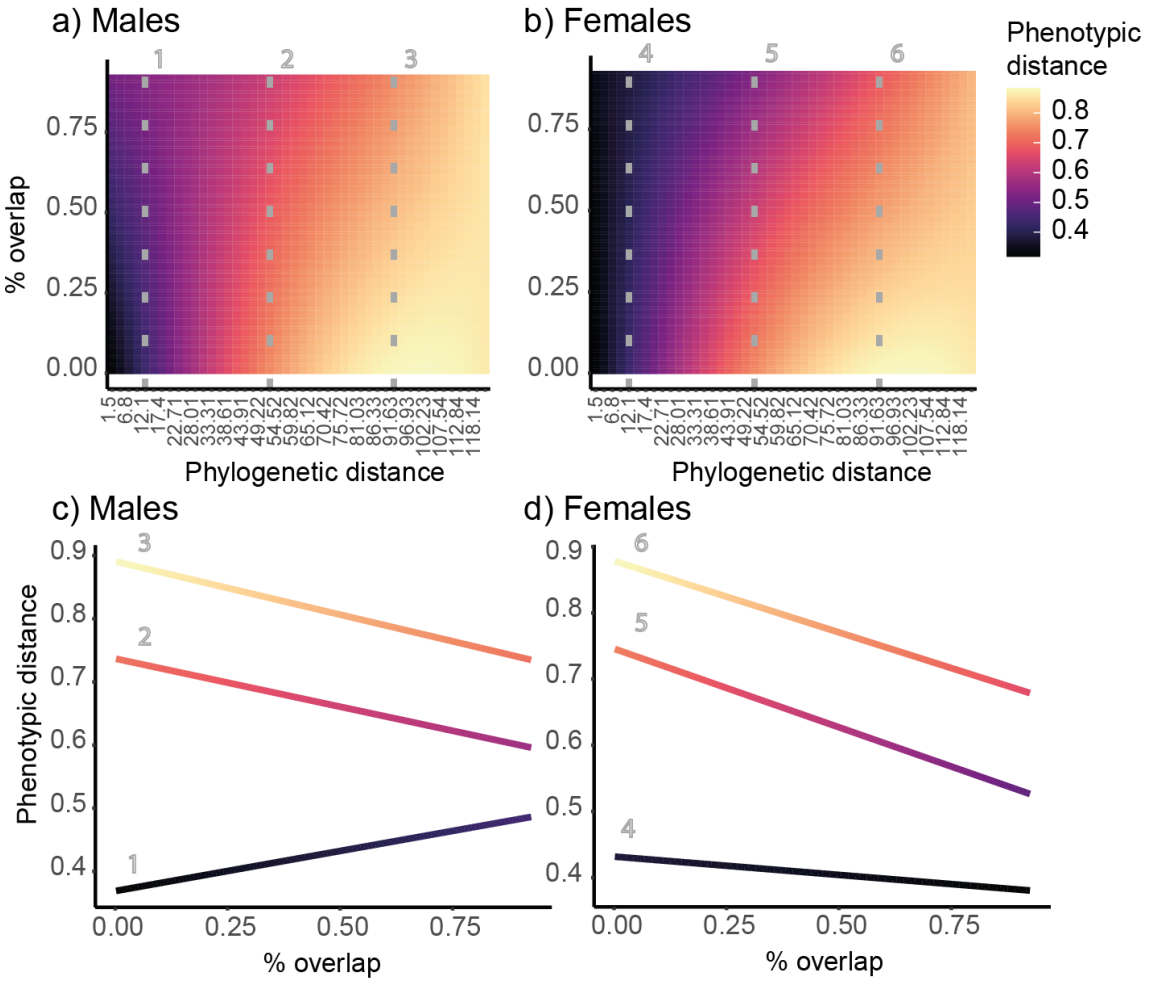

Figure 3. Complex interactions between the effect of phylogenetic distance and overlap on phenotypic distance. Heatmaps show predicted values of phenotypic distance in function of phylogenetic distance and percentage of overlap are shown for a) males and b) females. Example curves of predicted phenotypic distance in function of percentage of overlap for 3 arbitrarily chosen phylogenetic distances are shown for c) males and d) females. The 3 gray dotted lines for each heatmap show the phylogenetic distance chosen for the examples of curves of predicted phenotypic distance in function of percentage of

overlap - 1 and 4 are example curves for a phylogenetic distance of 12.1 Myr, 2 and 5 for 51.87 Myr and 3 and 6 for 91.63 Myr (sum of branch lengths).

To test for the overall effect of sympatry on phenotypic evolution, we then defined species that shared parts (≥20%) of their geographic range as sympatric, while all other species with non-overlapping distributions (<20%) were classified as allopatric. We then designed a permutation test to detect phenotypic convergence and divergence.

We first established a null distribution by 1) computing the residuals of the regression between the phenotypic and phylogenetic distances for each pair of phenotypic forms (noting that multiple phenotypic forms may exist within species), and then 2) permuting these residuals along the phylogeny. We then tested for each pair of forms whether they were more or less similar than expected from the null distribution, interpreted as the signature corresponding to convergence and divergence, respectively. We quantified the strength of convergence and divergence from the difference between the observations and the generated null distribution.

In males, out of the 29,402 pairs, we detected 256 convergent pairs and 193 divergent pairs, including 25 and 7 in sympatric pairs, respectively. In females, out of the 31,877 pairs, we detected 281 convergent pairs and 220 divergent pairs, including 55 and 7 in sympatric pairs, respectively. The relative number of detected convergence vs. divergence events was significantly greater in sympatric pairs as compared to allopatric pairs, for both males and females (χ^2= 4.54, p-value < 0.05 and χ^2= 25.37, p-value < 0.001, respectively). Overall, we found that sympatric pairs showed significant convergence in sympatry (p-value < 0.001 for males and females, Figure 4a, b). We also found similar convergence in sympatry when using a simulation-based approach for sympatric vs. allopatric pairs (Fig. S7 and S8),

and this trend was confirmed using different cut-offs for the degree of geographic overlap used to define sympatric pairs (see supplementary 4).

Focusing on detected convergence events, we did not find significant differences in convergence strength between sympatric and allopatric species pairs for males. In females, however, the convergence was significantly stronger in sympatric species pairs (Wilcoxon, $W = 7448$, p-value < 0.05). While we found convergence with overlap in males with the PLMMs, here we do not find a significantly higher convergence in sympatry vs allopatry for males using permutations. This stems from the significance threshold used for pairwise convergence/divergence assessment, which reduces the overall sympatric sample size compared to allopatric ones. Lower and less frequent convergence in males than females combined with reduced sample size due to the threshold may lead to the discrepancy of results for males between the two methods. The signal of convergence in sympatry found

in our global dataset suggests a differential effect of sympatry on the evolution of female compared to male phenotypes.

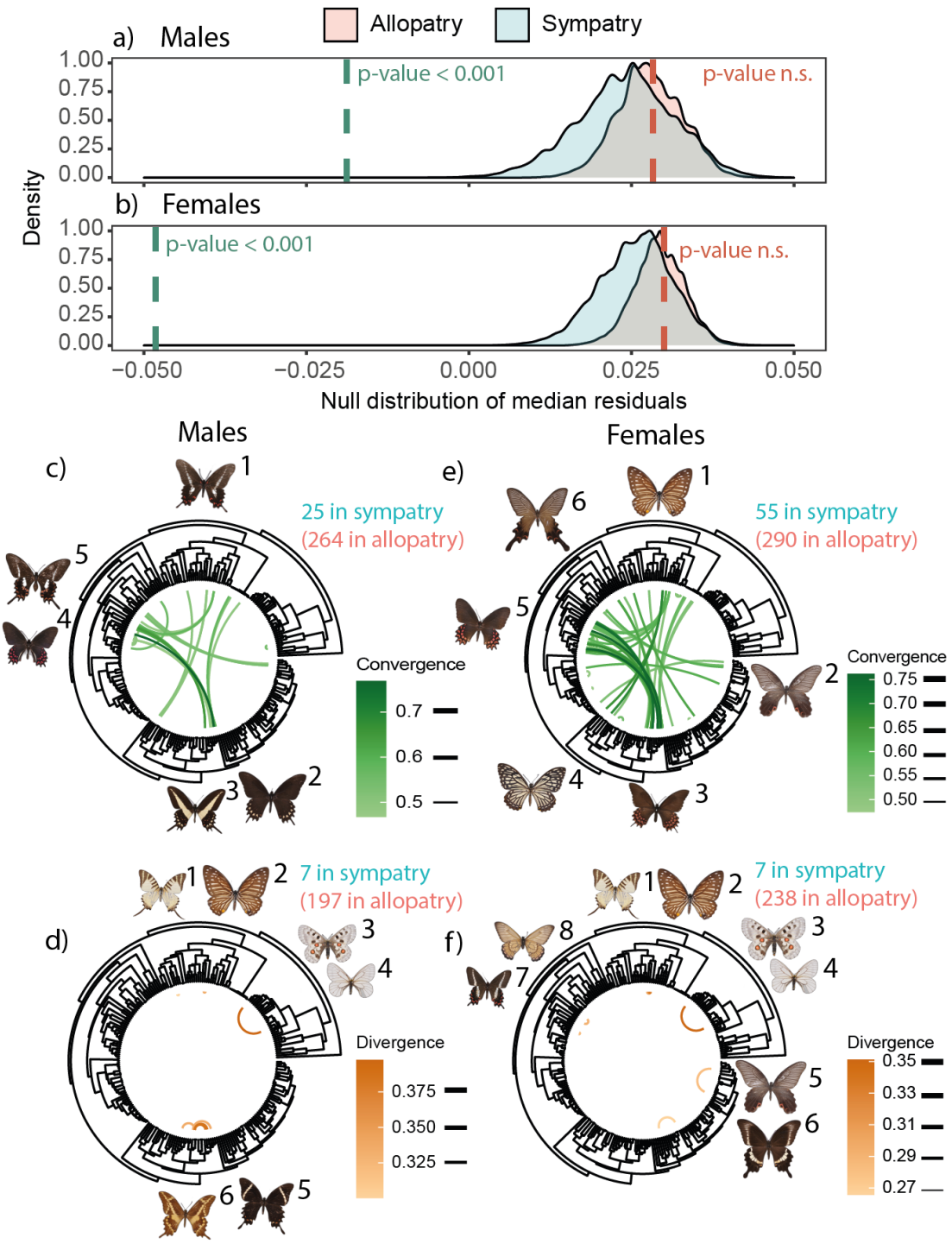

Figure 4. Detection of convergence and divergence events between sympatric and allopatric species pairs. Distribution of residual medians for (a) males and (b) females after 100,000 permutations for sympatric pairs in blue and allopatric pairs in pink. The dashed blue vertical line corresponds to the observed median for the residuals of pairs of sympatric species, and the dashed pink vertical line corresponds to the observed median for the pairs of allopatric species. (c-f) Phenotypic convergence and divergence associations between pairs represented on the swallowtail phylogenetic tree with male convergence (c) and divergence (d) between pairs, and female convergence (e) and divergence (f) between pairs. Example of pairs of convergent and divergent species are plotted along the phylogeny: c-1: *Mimoides lysithous,* c-2: *Papilio erostratus,* c-3: *Papilio hectorides,* c-4: *Parides photinus,* c-5: *Parides bunichus*. d-1: *Graphium antiphates,* d-2: *Graphium xenocles,* d-3: *Parnassius nomion,* d-4: *Parnassius stubbendorfii,* d-5: *Papilio pelaus,* d-6: *Papilio aristodemus*. e-1: *Graphium xenocles,* e-2: *Papilio protenor,* e-3: *Papilio erostratus,* e-4: *Papilio clytia,* e-5: *Parides photinus*. f-1 : *Graphium antiphates,* f-2: *Graphium xenocles,* f-3: *Parnassius nomion,* f-4: *Parnassius stubbendorfii,* f-5: *Papilio protenor,* f-6: *Papilio hipponous,* f-7: *Parides bunichus,* f-8: *Euryades corethrus*.

4. Evolution of sexual dimorphism driven by the evolution of female color pattern

To investigate the evolutionary divergence between male and female phenotypes within species, we estimated sexual dimorphism by computing the Euclidean distances in PCA space between the two sexes within each species. We also computed the raw contrasts for each pair of sister species – male/male distance and female/female distance in sister species – and examined the relative divergence of male phenotypes vs female phenotypes using a index based on the ratio of raw contrasts (r): when $r < 0$, female phenotypes diverged more than male phenotypes in the same amount of time (Figure 5). We investigated the distribution of the relative divergence of color pattern in each sex in dimorphic species to test for the relative effect of selection generated by reproductive interference on male phenotypes vs. ecological selection in female phenotypes on the evolution of sexual dimorphism. Dimorphic species were defined as species with dimorphism value above 0.3.

We found that most dimorphic sister-species pairs did indeed have a value of r lower than expected from no sex-specific signal (p-value < 0.05, Fig. S9), indicating that the sexual dimorphism in color pattern mostly stems from the divergence of female phenotype. This confirms that female phenotypes generally diverge more than male ones resulting in sexual dimorphism driven by ecological selection acting on females.

[Figure 5]

Figure 5. Evolution of sexual dimorphism in color pattern throughout the family Papilionidae: contribution of divergence of male phenotype vs. female phenotype in the evolution of sexual dimorphism was estimated by comparing sister-species. a) Swallowtail phylogeny showing sister-species with the level of dimorphism (gray scale) and ratio of raw contrasts: (male/male distance over female/female distance) - 1 for positive values and – (female/female distance over male/male distance) - 1 for negative value for sister species so that relative divergence becomes more pronounced as the value diverges from 0. b) Distribution of the ratio of raw contrasts for sister species with the median indicated. c) Male and female specimens are shown for the two-sister species presenting the highest ratio: (1) *Euryades corethrus* above and *Euryades duponchelii* below, and for the two sister species presenting the lowest ratio: (2) *Papilio meriones* above and *Papilio dardanus* below.

Discussion

Using a novel machine learning-based method, we quantified color pattern variation at the global geographic scale in Papilionidae butterflies and uncovered a general trend of convergence among sympatric species. This method is not specific to butterflies and could be trained to quantify variation in the visual signal in any biological model. By tuning the modifications made during the training, the method can be trained to quantify shape, color,

and pattern similarity in diverse organisms, opening new research avenues for the study of phenotypic diversification.

1. Uncovering discriminant features of visual signal

In this study, we provide the first quantitative analysis of butterfly color pattern variation at a large phylogenetic scale by comparing highly divergent and complex color patterns. The complexity and extreme diversity of butterfly color patterns has led previous studies to simplify the problem. Some authors have discretized the patterns (22, 29, 30) by focusing on a few characters at a time (31, 32). This approach comes at the cost of losing part of the signal as well as its global integration. Other approaches that rely on aligning wing images to the color patterns themselves allow for powerful comparisons based on entire color patterns (33). However, the alignment is limited to closely similar patterns and wing shapes, preventing analyses at large evolutionary scales. Here, we circumvent these issues with a new machine learning method to automatically quantify color pattern variation among distantly-related species in an unsupervised manner, independent of any pre-existing human classifications.

This unsupervised and automated method allows comparing multiple color patterns between very distantly-related species, which may differ in wing shape due to other selective constraints - see Le Roy et. al. (2019) (34) for a review. We therefore designed our method to neglect minor wing shape variation (see Sup. Mat Figure 7 showing the limited effect of wing shape variation on our estimates of distances between color patterns). Nevertheless, wing shape variation is not independent of color pattern and contributes to the visual signals perceived by conspecifics and predators. Linke et al. (2022) (35) showed that wild blue tits can learn to associate both color pattern and hindwing tails with escape ability, highlighting the importance of salient shape features such as wing tails in the generalization performed by visual observers. Although our method reduces the impact of minor variations in wing shape, it still retains information about prominent discriminative parts of the wings such as large tails (Fig. 2). Our method thus provides a relevant

quantification of color pattern variation within and between species for studying how selection exerted by visual predators and/or conspecifics shapes the diversification of wing color patterns.

Nevertheless, we only quantified visual signal similarity in the visible light range. Both butterflies and some of their avian predators are UV-sensitive, and some spots on Papilionidae wing patterns reflect UV wavelengths (e.g. in the white patches of *Papilio glaucus*, 36). Pigments or structural changes in the wing scale that result in UV reflectance generally also generate differences in the visible reflectance, but variations in UV reflectance may lead to conspicuous differences for UV-sensitive species that cannot be captured by our method. However, color pattern mimicry often involves convergence in the UV pattern as well (as shown in *Papilio polytes*, 37): ignoring the UV signal would thus not strongly alter the general trend of wing pattern convergence in Papilionidae.

2. Significant interactions between historical and ecological factors acting on trait diversification

By examining the evolution of color patterns throughout the whole Papilionidae family, we found increasing phenotypic similarity between species with increasing geographic overlap, when controlling for phylogenetic distance. Interestingly, we also found significant interactions between the effect of geographic overlap and phylogenetic distance. Phenotypic distance decreased more rapidly with geographic overlap at higher phylogenetic distances. Thus, the effect of sympatry on phenotypic convergence was greater in distantly related species, consistent with documented cases of Batesian mimicry in Papilionidae, with distantly-related pairs of species described as models and mimics. For closely-related males, the effect of the geographic overlap was actually opposite, with increased phenotypic divergence in closely-related species. This trend is consistent with the higher ratio of convergent pairs to divergent pairs in females as compared to males from sympatric species.

As expected, we found that sympatric pairs tended to be more closely related than allopatric pairs (Wilcoxon, males: p-value < 0.001, females: p-value < 0.001). This may stem from the history of species diversification, making recently diverged species more likely to occur in close geographic areas. Furthermore, recently diverged species may also tend to retain similar habitat affinities (phylogenetic niche conservatism) and species from a same clade thus frequently occupy similar geographic regions (38). Nevertheless, here we detected significant color pattern convergence in sympatric pairs despite this phylogenetic clustering, suggesting that local ecological processes might have fueled the evolution of traits. The use of geographic overlap as a proxy for sympatry between species has limitations, because species found in the same geographic location may still differ in phenology (39) or flight height (16), so that species sharing a similar geographic area do not necessarily occur in the same micro-habitat. Furthermore, we cannot directly assess whether trait convergence in sympatry is due to shared environmental conditions versus species interactions: alternative selective pressures due to climatic conditions for instance may promote similar color patterns in different species independently from a direct ecological interaction between these sympatric species. Nevertheless, the phylogenetic corrections account for trait similarities due to the combination of phylogenetic constraints or phylogenetic niche conservatism (40). The observed convergence is thus likely to be due to local selection exerted by predator behavior toward Papilionidae, where both defended and palatable species have been previously described to be involved in mimetic interactions (19). Further ecological studies on the local mimicry rings, including both Papilionidae and other butterfly families (e.g. Nymphalidae or Pieridae, 41), are now needed to formally conclude on the impact of predation on the evolution of convergent color patterns in sympatry.

3. The effect of sympatry on phenotypic convergence and divergence

Negative interactions in sympatry such as competition for resources and territory are often assumed to be the main driver of phenotypic evolution, in line with character displacement theory (2, 3). However, ecological interactions between sympatric species can also lead to

trait convergence (see Grether et al. 2009 (10) for a review). Here, we found an effect of sympatry on trait convergence between species, with a global trend of convergence in sympatric pairs and phenotypic similarity between species increasing with the degree of geographic overlap. Overall, of all the convergent pairs of sympatric species detected, 60% were previously reported in the literature as potential mimetic species (see the detailed list of all convergent and divergent sympatric pairs, and corresponding references in Supplementary Information). Despite our simplistic definition of sympatry, based on geographic range overlap without precise information on the shared microhabitat, our global-scale method was able to recover a large number of species pairs identified as mimetic in field studies. Indeed, female-limited Batesian mimicry has indeed long been documented in Papilionidae (20, 42), and our study recovers well-studied cases, such as the documented Batesian mimicry between Amazonian unpalatable *Parides* species, and their *Papilio* mimics – see for example the black pattern with conspicuous red spots in the unpalatable *Parides photinus* and its mimic *Papilio erostratus* (18) (Tab. S5). Interestingly, convergence was also detected within the genus *Parides*, for example between *Parides photinus* and *Parides montezuma*, supporting the hypothesis of a Müllerian mimicry ring within the genus *Parides* (43). Convergence was also found between the Southeast Asian species *Graphium xenocles* and a form of *Papilio clytia*, both displaying white coloration with contrasting black venation and an orange spot on the hindwing. While it is likely that both species are palatable, they are both mentioned as being mimics of unpalatable Danaine species and thus may belong to the same mimicry ring. Surprisingly, we also found convergence in males of sympatric species pairs. First, there is convergence between males from species where convergence was also detected in females, for example in *Parides photinus* and *Papilio erostratus* described above. The convergence in male phenotypes may be due to Batesian mimicry, but it may also stem from correlated color pattern evolution between males and females within the same species, caused by developmental constraints. In some species, convergence has been detected among palatable species, without any evidence for mimicry with defended species. Instead, these convergent species exhibit color patterns composed of contrasting black bars on a clear background and orange and blue spots on the hindwing near the tails, such as in the convergent *Papilio alexanor*

and *Iphiclides podalirius*. This type of color pattern, combined with hindwing tails, may help to deflect predators away from the vital body parts (44). Such convergence in color patterns could thus be driven by cognitive biases of predators and associated behavior, but does not imply mutualistic or parasitic interactions between butterfly species.

Such a strong convergence in coloration has also been documented in woodpeckers (45), where the effects of both abiotic factors and ecological interactions between sympatric species have been reported. Interestingly, a study focusing on color pattern variation in sympatric newt species revealed convergence in dorsal patterns involved in camouflage from predators and divergence in ventral colors, likely related to mate choice (8). In the Papilionidae, the dorsal wing color pattern is likely involved in interactions with both predators and mates. However, the global increase in convergence detected in sympatry suggests that selective pressures associated with predation outweigh reproductive character displacement in driving color pattern evolution, especially in females. This is illustrated in the PLMM analyses by the stronger effect of geographic overlap on phenotypic distance for more distantly related species pairs: in distantly related species, prezygotic isolation may be reinforced by multiple mating cues independent of color pattern, facilitating phenotypic convergence in sympatric species. In closely related species, however, the effect of increasing levels of geographic overlap is weaker, especially in males, where phenotypic divergence increases with the level of geographic overlap, consistent with the weaker signal of sympatric convergence detected in males compared to females. By revealing the important effect of sympatry on trait diversification at the macroevolutionary scale, our study suggests a correlation between local ecological processes and global trends of trait evolution.

4. Sex-specific selection pressures and the evolution of sexual dimorphism

In species where the level of parental investment in females is higher than in males, we expect increased selection on female choosiness (46) and visual discrimination abilities (as observed in mimetic butterflies such as *Heliconius erato*, 47). In Papilionidae, males

attempt to mate several times during their adult life, while females mate very early in life and remating is generally rare. Females spend most of their adult life searching for and laying eggs on suitable host plants. Therefore, the cost of mating with a heterospecific is probably higher for females than for males, leading to the evolution of female preference for non-mimetic males (24). In Papilionidae, we found significantly more and stronger cases of convergence in sympatry compared to allopatry for females, but not for males. This is consistent with the previously described high prevalence of female-limited Batesian mimicry in Papilionidae. For instance, in the genus *Papilio*, approximately 25% of species have mimetic females, while males generally display the ancestral color pattern (48). Overall, we found that female color patterns diverged more than male color patterns in sister species. Color pattern dimorphism appears to be driven mostly by natural selection acting on female phenotypes (22), but may also be favored by female preferences for males displaying ancestral phenotypes, limiting heterospecific interactions.

Our method is, however, likely to fail for detecting some cases of sexual dimorphism driven by divergence in male phenotypes in sympatric species, implying UV reflectance: some sister species with partially overlapping geographic distributions, such as *Iphiclides podalirius* and *I. feisthamelii*, have very similar color patterns, but males differ in UV reflectance of some patches, while no such differences were detected in females (49). This divergence in male coloration could be due to selection caused by reproductive interference in these parapatric sister species. Our method aims at detecting sexual dimorphism generated by ecological interactions in sister species. This focus on sister species discards the effect of phylogenetic correction bias but prevents the detection of ancestral sexual dimorphism. For example, the striking sexual dimorphism observed in the genus *Ornithoptera*, in which males display brightly saturated green, blue, and orange colors while females are more melanic and cryptic, is not detected in our test as resulting from male divergence. Such ancestral sexual dimorphism that might stem from sexual selection acting on males is likely to be independent of reproductive interference between sympatric species.

Conclusion

The method described here allowed to uncover global trends on a large phylogenetic scale, and is likely to stimulate new studies on the diversification of visual traits that have been difficult to compare so far. Our study of color pattern diversification in Papilionidae suggests a potentially important role for local ecological processes in sympatry in macroevolutionary patterns of trait diversification.

Materials and Methods

1. Sampling and standardized photographs

We sampled specimens from the collection of the National Museum of Natural History (Paris) to cover most of the described Papilionidae species. We relied on the latest taxonomic reviews (50) and published dated species-level phylogeny (25) to scan the collections for the described species. We sampled 337 out of the 382 species shown in the phylogeny of Allio et al. (2021) (25). For each species, we selected two males and two females, whenever specimens were available. 17 species presented multiple phenotypic forms (2 in average for a total of 32 forms, which we sampled). Our sample thus consisted of 1,358 individuals, including 774 males (329 species) and 592 females (273 species; note that females are very rarely collected so that females were lacking for many species). The dorsal and ventral side of the sampled individuals were photographed under controlled LED light using a Nikon D90 (Camera lens: AF-S Micro Nikkor 60 mm 1:2.8G ED) under standardized conditions. For the statistical analysis, we retained only species for which we had at least one male and one female, resulting in a final sample size of 267 species, 292 subspecies, and 296 unique phenotypic forms for males and 313 for females.

2. Wing segmentation

For each picture, the four wings were first digitally separated from the body and from the background of the pictures using a combination of machine learning-based segmentation and traditional image processing. The machine learning-based segmentation was

performed using a Mask-RCNN model, which learns to classify each pixel on the picture either as belonging to the region of interest or not and generates a segmentation mask based on this classification. First, a training database of 371 pictures was constituted, by manually cropping the four wings on each picture. A Mask-RCNN model was trained on two thirds of this database (247 pictures) using the PixelLib python library. The rest of the pictures were used to evaluate the segmentation using the intersection over union of masks metric (IOU) with a threshold of 0.9 to consider the prediction a true positive. The IOU with this threshold of 0.9 was 0.85, *i.e.* for 85% of the predicted masks, the mask was 90% or more in agreement with manual cropping. As the produced masks often kept a few background pixels at the edges of the wings, a supplementary step was added, using traditional image processing. After this post-processing step, the IOU with a threshold of 0.9 went up to 0.98, meaning that 98% of the predicted masks were consistent with 90% of the manually cropped images. However, in a few pictures, wing parts whose coloration closely matched the background color had to be manually eliminated. Finally, we obtained the masks for the four wings for each of the 2,716 standardized pictures, and generated images with a single wing on a white background.

3. SimCLR training and evaluation

Wing coloration is a complex character allowing to accurately measure the strength of convergence and divergence despite phylogenetic constraints, but accurately quantifying its variation is challenging. Most studies rely on character discretization (22, 29) or focus on a few elements of the color pattern (31, 32), and quantitative methods are scarce and require an alignment step of the wings to allow direct comparison (30, 33, 51), preventing comparisons at large evolutionary scales due to important differences in wing shape. Recently developed convolutional neural networks specialized for image analysis can perform image classification, segmentation of regions of interest, or even similarity learning, and have been successfully applied to wing color patterns in Lepidoptera (52–54). However, these methods rely on supervised learning, which requires labels such as species classification, and thus introduces strong biases in phenotypic comparisons. Here, we develop an unsupervised machine learning-based method to quantify color pattern

similarity independent of phylogenetic information, allowing us to estimate the relative effect of species interactions in sympatry vs. shared ancestry.

The SimCLR method

The similarity of color patterns was quantified using the new unsupervised deep contrastive metric learning method SimCLR (27). This method is designed to provide a distance metric between images and can produce the images' coordinates in a reduced dimensional space. Traditionally, supervised contrastive learning needs labels for the images to determine whether pairs of images are considered similar (positive pair, belonging to the same class) or dissimilar (negative pair, belonging to different classes), allowing the neural network to learn features making images similar or dissimilar by modifying the weight of neurons during training to obtain a smaller distance for positive pairs and a larger one for negative pairs. SimCLR does not require such labeling of the pairs but considers each image as their own class, using modified versions of the image to perform comparisons. These modifications, called image augmentations, include cropping, rotating, inverting, and color jittering, and are performed at a rate set by the user. Augmentations of the same initial image are considered as positive pairs, while augmentations of different images are considered as negative pairs. Metric learning is computed by minimizing a loss function (Normalized Temperature Cross Entropy Loss or NT-Xent loss). The loss is minimized by adjusting the weights of the neural network during learning to increase the cosine similarity within positive pairs and decrease it for negative pairs. The cosine similarity between two vectors u and v is defined as $sim(u,v) = \frac{u \cdot v}{\|u\|\|v\|}$. For a positive pair (i,j) in a batch of size N, the NT-Xent loss is defined as:

$$l_{i,j} = -\log \frac{exp\left(\frac{sim(z_i, z_j)}{\tau}\right)}{\sum_{k=1}^{2N} 1_{[k \neq i]} exp\left(\frac{sim(z_i, z_k)}{\tau}\right)}$$

where $1_{k \neq i} \in \{0,1\}$ is an indicator function evaluating to 1 if $k \neq i$ and τ denotes a temperature parameter.

Image augmentations

The choice of augmentations is crucial as it determines the features to which the method will be invariant. Thus, to study the variation in color pattern, we applied randomized cropping with a minimum size of 50% of the original image (invariance to size); rotation, horizontal and vertical flipping each with a probability of 50% (invariance to left/right/up/down or orientation); conversion to grayscale with a probability of 20% (invariance to pattern without taking color into account); and finally, no color jittering of any kind to account for the color hue, brightness, and saturation of the actual wings.

Network's architecture

Based on Chen et al. (2020) (27), we chose a ResNet50 (residual neural network with 50 convolutional layers) as the backbone of the network and then replaced the classification head with a SimCLR multilinear projection head. The Resnet50 backbone was initialized as pre-trained on the large ImageNet database to increase performance and compensate for the limited number of images in our own dataset. Three hyperparameters were optimized: the batch size, the number of training epochs and the temperature parameter. A grid search was performed to find the optimal parameter values, e.g., the parameters that lead to the best performance during evaluation.

Evaluation & training

After the training phase, the output of the last convolutional layer was used as the vector representation of the images, since it retains more information than the output of the multilinear projection head. The evaluation was based on a pre-training task aimed at classifying the vector representation of the images into different categories. To obtain

categories from our images without *a priori*, we classified our images into 16 clusters, using the HDBSCAN clustering method (55) on the vector representation obtained with a classical pre-trained convolutional neural network, VGG16 (56). The cluster labels were then used as class labels during classification. The backbone of the SimCLR method was also pre-trained on the same dataset. Unlike the SimCLR method, the vector representations from the VGG16 network did not provide a relative metric space. Finally, to assess if the pre-training biases the evaluation, we compared performance of the pre-trained-only method, the pre-trained and fine-tuned during classification method, and the SimCLR trained method. A more detailed discussion of the evaluation of the method can be found in Supplementary Information 1. The method was implemented in Python, mainly using the Pytorch library for machine learning, and the Lightly library for SimCLR related augmentations, backbone, and loss function. Finally, the method was trained with a batch size of 128, several training epochs of 300 and a temperature of 0.5, obtaining an f1 score 1.15% higher than the pre-trained-only model, a kappa score 1.19% higher, and a mean accuracy of 95% (84% for the pre-trained-only method). A PCA then allowed the dimensions of the representation vectors to be reduced from 2,048 to 20, while retaining approximately 80% of the variance.

4. Explainability and quality control

To identify the features of the images used for discrimination by the SimCLR method, we generated a gradient-weighted class activation mapping (Grad-CAM, 56) for the input images. This allows pinpointing pixels of the image generating the highest activation in the convolutional layers. To check for the reliability of the image embeddings, the distances between all image pairs were calculated. We then compared the distribution of intraspecific and interspecific pairwise distances to test whether images of butterflies belonging to the same species had a lower phenotypic distance than interspecific image pairs.

5. Species geographical range

For 125 out of the 267 species studied here, the geographical range was retrieved from the IUCN Red List Spatial Data & Mapping Resources (58) or from the Map Of Life project (59) using different datasets. When the geographic range could not be retrieved from these

sources because it was missing or incomplete (101 species) the geographic range was estimated from GBIF occurrences by generating a convex alpha hull with a buffer distance around the occurrences. GBIF occurrence data were downloaded and cleaned using the rgbif and CoordinateCleaner R packages (60) and species with fewer than 30 occurrences were discarded. For 45 species, geographical range could not be retrieved because of missing or too few GBIF data. These species were excluded from the analysis. Pairwise species overlap was calculated using the Jaccard index (area of the intersection of the ranges over area of the union of ranges). A pair of species was considered sympatric if their ranges overlapped by 20% or more, and allopatric if they did not.

6. Detection and quantification of convergence and divergence using permutation tests

Convergent and divergent pairs

To detect and quantify convergence and divergence, we designed a permutation-based method similar to that described (61). Because many species have multiple forms, we computed the mean form phenotypes as the mean of the vectors of all specimens belonging to the same form. To assess the degree of phenotypic similarity independent of phylogenetic proximity, we performed a linear regression between pairwise phenotypic distances (computed as the Euclidean distance between the phenotypes) and pairwise phylogenetic distances. Negative residuals represent pairs for which the phenotypic distance is smaller than expected given the phylogenetic distance between species, which may indicate putative phenotypic convergence. Conversely, positive residuals indicate greater phenotypic divergence than expected from the phylogenetic distance. To determine the degree of wing color pattern convergence in the swallowtail phylogeny, we tested whether the studied pairs were more phenotypically similar than expected at random, while controlling for phylogenetic distances. We thus compared the residuals obtained above to a null distribution of residuals. The null distribution was generated by permuting the residuals associated with each species throughout the phylogeny. The permutations were performed using the Lapointe-Garland permutation method, in which pairs of trait values are exchanged with probability inversely proportional to their phylogenetic distance (62).

This permutation method allows correcting for non-independent and identically distributed data, therefore removing phylogenetically induced false positives and accounting for unbalanced phenotypic distribution. Residuals were permuted 100,000 times over the 29,402 possible pairs for males and 31,877 for females. The *p*-value associated with phenotypic convergence for each pair was calculated as the proportion of permutations where the observed residuals was lower than the permuted residuals. We fixed the *p*-value at 1% and thus deemed convergence between a pair of phenotypic forms significant if 99% of the permuted residuals for this pair were higher than the observed residual. Conversely, to test for phenotypic divergence events, we performed the same permutation test but computed for each pair the proportion of permutations where the observed residual was greater than the permuted residual to obtain a *p*-value and fixed the *p*-value at 1%. Convergence and divergence strength were then quantified as the deviation of the observed residuals from the median computed in the generated null distribution.

Sympatry and allopatry

To assess the overall convergence or divergence of sympatric pairs, we computed the median of the permuted residuals for sympatric species pairs for each permutation. A *p*-value was computed by counting the number of permutations in which the permuted median for sympatric pairs was greater than the observed median for sympatric pairs. To compare sympatry with allopatry, the same was done for allopatric pairs.

7. Phylogenetic Linear Mixed Models (PLMM)

To assess the effect of percent overlap on phenotypic distance while controlling for phylogenetic distance, we fitted PLMM at the species level with phenotypic distance as the response variable and percent overlap, phylogenetic distance and phylogenetic distance squared and their interactions as predictors. Additionally to these fixed effects, we fitted the identity of the species 1 of the pair and the species 2 of the pair as random effects while

linking each to the phylogeny, in order to take into account the non-independence of species in pairs in a standardized phylogenetic comparative framework. The quadratic term of phylogenetic distance squared was added to account for the nonlinearity of accumulation of phenotypic distance with respect to phylogenetic distance. Predictor variables were standardized prior to fitting, and PLMM models were fitted using the R package phyr (63).

8. Dimorphism in sister species

To assess which sex phenotype is driving the evolution of dimorphism, we determined which sex has diverged phenotypically more than the other in the same amount of time for each pair of sister species. We computed the Euclidean distance between the male phenotypic coordinates and the Euclidean distance between the female phenotypic coordinates of the two species in each pair, which correspond to the raw contrasts for males and females, respectively. To compare the two, we computed the ratio of contrasts $\frac{Euclidean(male_{species\,1}, male_{species\,2})}{Euclidean(female_{species\,1}, female_{species\,2})} = \frac{raw\ contrast\ males}{raw\ contrast\ females}$ and the inverse $\frac{raw\ contrast\ females}{raw\ contrast\ males}$ for each sister species. Because these quantities will be bounded by 0 in one direction and unbounded in the positive values, we arbitrarily defined female biased divergence as negative and male biased divergence as positive in the following way: When females have diverged more we use $-(\frac{raw\ contrast\ females}{raw\ contrast\ males} - 1)$ and when male diverged more$(\frac{raw\ contrast\ males}{raw\ contrast\ females}) - 1$. This allows us to compare divergence values for male biased and female biased phenotypic divergence, so that relative divergence becomes more pronounced as values diverge from 0 with female biased divergence in the negatives and male biased divergence in the positives. We assessed the level of dimorphism in each species by taking the Euclidean distance between males' phenotypic coordinates and females phenotypic coordinates.

To test whether this index is lower than expected if males and females diverged equally, we permuted phenotypes between sexes with equal probability, for each species. This

approach thus erases any sex-specific signal. We performed 10,000 permutations and recorded the median of the index for each permutation. We then compared the observed median of the actual index in our sister species pairs, with the distribution obtained and calculated the p-value as the frequency of permutations with a lower median index.

Data availability

All the code necessary for the python machine learning training and the R analysis, as well as the phenotypic coordinates in the morpho space are provided in the following Github: https://github.com/AgathePuissant/SimCLR_phenotypic_convergence. Specimen photographs are available upon request.

Acknowledgments

This work was funded by UMR 7205 from Muséum National d'Histoire Naturelle. We thank the intensive computing platform « Plateforme de Calcul Intensif et Algorithmique PCIA, *Muséum national d'histoire naturelle, Centre national de la recherche scientifique, UAR 2700 2AD, CP 26, 57 rue Cuvier, F-75231 Paris Cedex 05, France* » for allowing model training. F.L.C. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project GAIA, agreement no. 851188). We thank Vincent Debat for commentaries on the manuscript, Maël Doré for help with the permutation analysis, Julien Clavel for help in interpretation of PLMM analysis and Jasmine L. Hardie for helpful comments on the dimorphism analysis.

References

1. J. Cavender-Bares, D. D. Ackerly, D. A. Baum, F. A. Bazzaz, Phylogenetic Overdispersion in Floridian Oak Communities. *The American Naturalist* **163**, 823–843 (2004).

2. D. W. Pfennig, K. S. Pfennig, Character Displacement and the Origins of Diversity. *The American Naturalist* **176**, S26–S44 (2010).

3. K. S. Pfennig, D. W. Pfennig, Character Displacement: Ecological and Reproductive Responses to a Common Evolutionary Problem. *The Quarterly Review of Biology* **84**, 253–276 (2009).

4. J. P. Drury, K. W. Okamoto, C. N. Anderson, G. F. Grether, Reproductive interference explains persistence of aggression between species. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20142256 (2015).

5. Y. Huang, X. Wang, X. Yang, J. Jiang, J. Hu, Unveiling the roles of interspecific competition and local adaptation in phenotypic differentiation of parapatric frogs. *Current Zoology* **66**, 383–392 (2020).

6. D. C. Adams, M. E. West, M. L. Collyer, Location-specific sympatric morphological divergence as a possible response to species interactions in West Virginia Plethodon salamander communities. *Journal of Animal Ecology* **76**, 289–295 (2007).

7. N. Chazot, *et al.*, Mutualistic Mimicry and Filtering by Altitude Shape the Structure of Andean Butterfly Communities. *The American Naturalist* **183**, 26–39 (2014).

8. T. de Solan, *et al.*, A lot of convergence, a bit of divergence: Environment and interspecific interactions shape body colour patterns in Lissotriton newts. *Journal of Evolutionary Biology* **35**, 575–588 (2022).

9. J. B. Losos, Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* **11**, 995–1003 (2008).

10. G. F. Grether, N. Losin, C. N. Anderson, K. Okamoto, The role of interspecific interference competition in character displacement and the evolution of competitor recognition. *Biological Reviews* **84**, 617–635 (2009).

11. H. L. Kenyon, P. R. Martin, Experimental test of selection against hybridization as a driver of avian signal divergence. *Journal of Evolutionary Biology* **35**, 1087–1098 (2022).

12. J. C. Hinojosa, *et al.*, Rapid colour shift by reproductive character displacement in Cupido butterflies. *Molecular Ecology* **29**, 4942–4955 (2020).

13. E. M. Lemmon, Diversification of Conspecific Signals in Sympatry: Geographic Overlap Drives Multidimensional Reproductive Character Displacement in Frogs. *Evolution* **63**, 1155–1170 (2009).

14. S. Benitez-Vieyra, N. H. de Ibarra, A. M. Wertlen, A. A. Cocucci, How to look like a mallow: evidence of floral mimicry between Turneraceae and Malvaceae. *Proceedings of the Royal Society B: Biological Sciences* **274**, 2239–2248 (2007).

15. D. D. Kapan, Three-butterfly system provides a field test of müllerian mimicry. *Nature* **409**, 338–340 (2001).

16. K. R. Willmott, J. C. Robinson Willmott, M. Elias, C. D. Jiggins, Maintaining mimicry diversity: optimal warning colour patterns differ among microhabitats in Amazonian clearwing butterflies. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20170744 (2017).

17. M. Katoh, H. Tatsuta, K. Tsuji, Rapid evolution of a Batesian mimicry trait in a butterfly responding to arrival of a new model. *Sci Rep* **7**, 6369 (2017).

18. R. M. Merrill, *et al.*, The diversification of Heliconius butterflies: what have we learned in 150 years? *Journal of Evolutionary Biology* **28**, 1417–1438 (2015).

19. K. Kunte, The Diversity and Evolution of Batesian Mimicry in Papilio Swallowtail Butterflies. *Evolution* **63**, 2707–2716 (2009).

20. C. Clarke, P. M. Sheppard, The genetics of the mimetic butterfly Papilio polytes L. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **263**, 431–458 (1997).

21. R. M. Merrill, *et al.*, Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B: Biological Sciences* **279**, 4907–4913 (2012).

22. K. Kunte, Mimetic butterflies support Wallace's model of sexual dimorphism. *Proceedings of the Royal Society B: Biological Sciences* **275**, 1617–1624 (2008).

23. S. Su, M. Lim, K. Kunte, Prey from the eyes of predators: Color discriminability of aposematic and mimetic butterflies from an avian visual perspective. *Evolution* **69**, 2985–2994 (2015).

24. L. Maisonneuve, C. Smadi, V. Llaurens, Evolutionary origins of sexual dimorphism: Lessons from female-limited mimicry in butterflies. *Evolution* **76**, 2404–2423 (2022).

25. R. Allio, *et al.*, Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. *Nat Commun* **12**, 354 (2021).

26. H. F. Nijhout, Elements of butterfly wing patterns. *Journal of Experimental Zoology* **291**, 213–225 (2001).

27. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A Simple Framework for Contrastive Learning of Visual Representations in *Proceedings of the 37th International Conference on Machine Learning*, (PMLR, 2020), pp. 1597–1607.

28. J. Clavel, G. Escarguel, G. Merceron, mvmorph: an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution* **6**, 1311–1319 (2015).

29. H. M. Hines, P. H. Williams, Mimetic colour pattern evolution in the highly polymorphic Bombus trifasciatus (Hymenoptera: Apidae) species complex and its comimics. *Zoological Journal of the Linnean Society* **166**, 805–826 (2012).

30. S. M. Van Belleghem, *et al.*, patternize: An R package for quantifying colour pattern variation. *Methods in Ecology and Evolution* **9**, 390–398 (2018).

31. J. C. Oliver, K. A. Robertson, A. Monteiro, Accommodating natural and sexual selection in butterfly wing pattern evolution. *Proceedings of the Royal Society B: Biological Sciences* **276**, 2369–2375 (2009).

32. E. Ortiz-Acevedo, J. P. Gomez, M. Espeland, E. F. A. Toussaint, K. R. Willmott, The roles of wing color pattern and geography in the evolution of Neotropical Preponini butterflies. *Ecology and Evolution* **10**, 12801–12816 (2020).

33. Y. Le Poul, *et al.*, Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature communications* **5**, 1–8 (2014).

34. C. Le Roy, V. Debat, V. Llaurens, Adaptive evolution of butterfly wing shape: from morphology to behaviour. *Biological Reviews* **94**, 1261–1281 (2019).

35. D. Linke, M. Elias, I. Klečková, J. Mappes, P. Matos-Maraví, Shape of Evasive Prey Can Be an Important Cue That Triggers Learning in Avian Predators. *Frontiers in Ecology and Evolution* **10** (2022).

36. M. L. Aardema, Ultraviolet Coloration in Tiger Swallowtail Butterflies (Papilio glaucus Group, Papilionidae) with a Method for Objectively Quantifying Adult Butterfly Wing Wear. *lepi* **69**, 58–62 (2015).

37. S. Yoda, *et al.*, Genetic switch in UV response of mimicry-related pale-yellow colors in Batesian mimic butterfly, Papilio polytes. *Science Advances* **7**, eabd6475 (2021).

38. J. Cavender-Bares, K. H. Kozak, P. V. A. Fine, S. W. Kembel, The merging of community ecology and phylogenetic biology. *Ecology Letters* **12**, 693–715 (2009).

39. C. Le Roy, *et al.*, Convergent morphology and divergent phenology promote the coexistence of Morpho butterfly species. *Nat Commun* **12**, 7248 (2021).

40. M. Westoby, M. R. Leishman, J. M. Lord, On Misinterpreting the `Phylogenetic Correction'. *Journal of Ecology* **83**, 531–534 (1995).

41. J. Joshi, A. Prakash, K. Kunte, Evolutionary Assembly of Communities in Butterfly Mimicry Rings. *The American Naturalist* **189**, E58–E76 (2017).

42. C. A. Clarke, P. M. Sheppard, I. W. B. Thornton, The genetics of the mimetic butterfly Papilio memnon L. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **254**, 37–89 (1997).

43. A. B. B. de Morais, K. S. Brown, Larval foodplant and other effects on troidine guild composition (Papilionidae) in southeastern Brazil. *The Journal of Research on the Lepidoptera* **30**, 19–37 (1991).

44. A. Chotard, *et al.*, Evidence of attack deflection suggests adaptive evolution of wing tails in butterflies. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220562 (2022).

45. E. T. Miller, G. M. Leighton, B. G. Freeman, A. C. Lees, R. A. Ligon, Ecological and geographical overlap drive plumage evolution and mimicry in woodpeckers. *Nat Commun* **10**, 1602 (2019).

46. H. Kokko, R. A. Johnstone, Why is mutual mate choice not the norm? Operational sex ratios, sex roles and the evolution of sexually dimorphic and monomorphic signalling. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **357**, 319–330 (2002).

47. K. J. McCulloch, D. Osorio, A. D. Briscoe, Sexual dimorphism in the compound eye of Heliconius erato: a nymphalid butterfly with at least five spectral classes of photoreceptor. *Journal of Experimental Biology* **219**, 2377–2387 (2016).

48. K. Kunte, Female-limited mimetic polymorphism: a review of theories and a critique of sexual selection as balancing selection. *Animal Behaviour* **78**, 1029–1036 (2009).

49. A. Gaunet, *et al.*, Two consecutive Wolbachia-mediated mitochondrial introgressions obscure taxonomy in Palearctic swallowtail butterflies (Lepidoptera, Papilionidae). *Zoologica Scripta* **48**, 507–519 (2019).

50. M. Nakae, Y. Nishiyama, A. Cotton, *Papilionidae of the World* (2021).

51. D. N. Basu, V. Bhaumik, K. Kunte, The tempo and mode of character evolution in the assembly of mimetic communities. *Proceedings of the National Academy of Sciences* **120**, e2203724120 (2023).

52. J. F. H. Cuthill, N. Guttenberg, S. Ledger, R. Crowther, B. Huertas, Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. *Science advances* **5**, eaaw4967 (2019).

53. B. D. Ezray, D. C. Wham, C. E. Hill, H. M. Hines, Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum. *Proceedings of the Royal Society B: Biological Sciences* **286**, 20191501 (2019).

54. S. Wu, *et al.*, Artificial intelligence reveals environmental constraints on colour diversity in insects. *Nat Commun* **10**, 4554 (2019).

55. L. McInnes, J. Healy, S. Astels, hdbscan: Hierarchical density based clustering. *JOSS* **2**, 205 (2017).

56. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition (2015) https:/doi.org/10.48550/arXiv.1409.1556 (June 27, 2022).

57. R. R. Selvaraju, *et al.*, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int J Comput Vis* **128**, 336–359 (2020).

58. , The IUCN Red List of Threatened Species. *IUCN Red List of Threatened Species* (June 30, 2022).

59. W. Jetz, J. M. McPherson, R. P. Guralnick, Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution* **27**, 151–159 (2012).

60. A. Zizka, *et al.*, CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**, 744–751 (2019).

61. C. S. Pinna, *et al.*, Mimicry can drive convergence in structural and light transmission features of transparent wings in Lepidoptera. *eLife* **10**, e69080 (2021).

62. F.-J. Lapointe, T. Garland, A Generalized Permutation Model for the Analysis of Cross-Species Data. *J. Classification* **18**, 109–127 (2001).

63. D. Li, R. Dinnage, L. A. Nell, M. R. Helmus, A. R. Ives, phyr: An r package for phylogenetic species-distribution modelling in ecological communities. *Methods in Ecology and Evolution* **11**, 1455–1463 (2020).

Supporting Information

1. Evaluation of the method. Unsupervised deep learning is a powerful tool that allowed us to objectively quantify image similarity, without introducing biases due to human labelling – e.g., systematic classification such as species. However, we had to find a way to evaluate the neural network's learning. The way to evaluate SimCLR's learning is to design a downstream pretext task such as classification using only the embeddings learned from the images, allowing to assess if the information contained in the embeddings is enough to classify correctly. In our case, because of the few images per species and sex (from 1 to 5), we could not use species as labels for such a pretext task, because it led to systematic overfitting of the classifier. Instead, we found a workaround using clustering to create pseudo labels for the classification. Using a pretrained network, we clustered images into 16 groups, representing broad phenotypic groups, which we used as labels for classification. The pre-training being on the same database that our SimCLR method (ImageNet), the evaluation is not completely independent from the training. However, we do not dispose of a ground truth or enough images and so this workaround was the best way to evaluate the model. Moreover, we took care of comparing SimCLR only pre-trained with SimCLR further trained on our database, which showed improvement in performance, showing that we effectively learned information specific to our data.

To assess the ability of the learnt embeddings to contain information about independent taxonomic labels, we went up one taxonomic level and trained a classifier on the embeddings to classify species in the right genus. This way there is enough training data in each class to train a classifier, even if phenotypically diverse genera may pose a challenge to classification. We retrieve a mean accuracy of 70%, with a minimum of 45% for the *Graphium* genus (a highly diverse genus) and a maximum of 100% for 17 genus out of the 29. We retrieve a f1 score of 0.95 and a kappa score of 0.94, which indicates a good performance of the classifier. Below is a table detailing accuracy for each class (table S2)

	Pre-trained-only	Pre-trained and fine-tuned	SimCLR trained
Mean accuracy	84%	87%	95%
F1 score	0.84	0.88	0.97
Kappa score	0.81	0.86	0.96

Table. S1. Evaluation metrics for the classification network only pretrained on ImageNet, fine tuned on our butterfly pictures, and trained with SimCLR on our butterfly pictures. Mean accuracy is the mean percentage of correct prediction over all classes. F1 score is a measure between 0 and 1 with 1 being the value for a network that perfectly predicts classes, that takes into account both precision and recall, i.e. number of true positives and false positives. Kappa score is a parameter ranging from 0 to 1, estimating reliability while correcting for the number of classes. Kappa=1 is obtained when a network perfectly predicts classes.

Class	**Accuracy(%)**
Parides	100
Mimoides	100
Battus	93.1
Hypermnestra	100

Iphiclides	75
Archon	100
Sericinus	100
Troides	100
Teinopalpus	100
Ornithoptera	84.9
Protesilaus	100
Pharmacophagus	100
Trogonoptera	100
Baronia	100
Bhutanitis	100
Allancastria	100
Meandrusa	94.1
Lamproptera	93.8
Graphium	45

Parnassius	90.3
Eurytides	81.9
Luehdorfia	100
Protographium	87.5
Pachliopta	100
Papilio	100
Atrophaneura	66.7
Zerynthia	100
Cressida	88.4
Euryades	85.7

Table S2. Accuracy of classification of phenotypes into each genus, learned from SimCLR embeddings only.

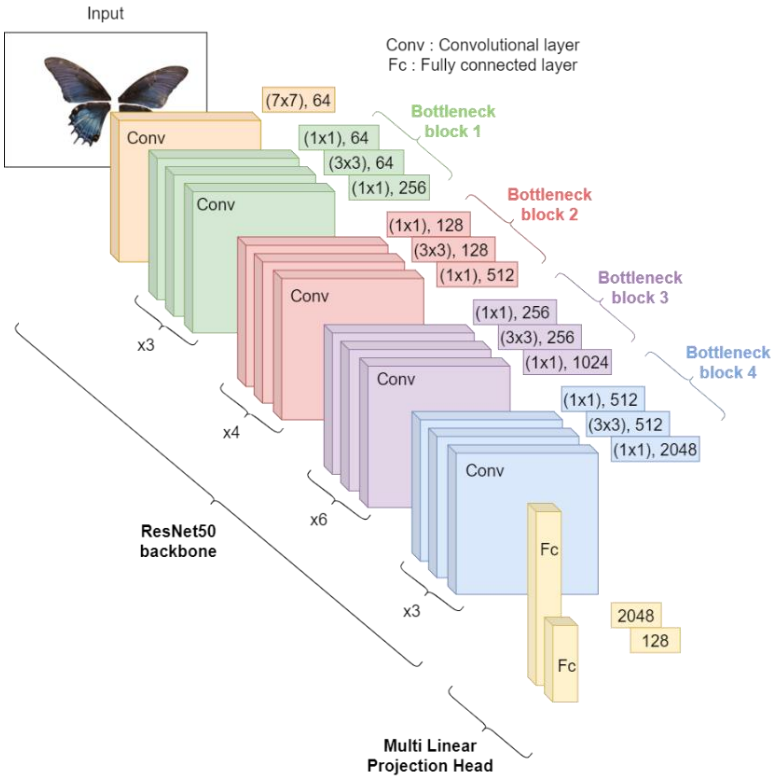

Fig. S1. Network architecture made of a Resnet50 convolutional neural network backbone and a multilinear projection head.

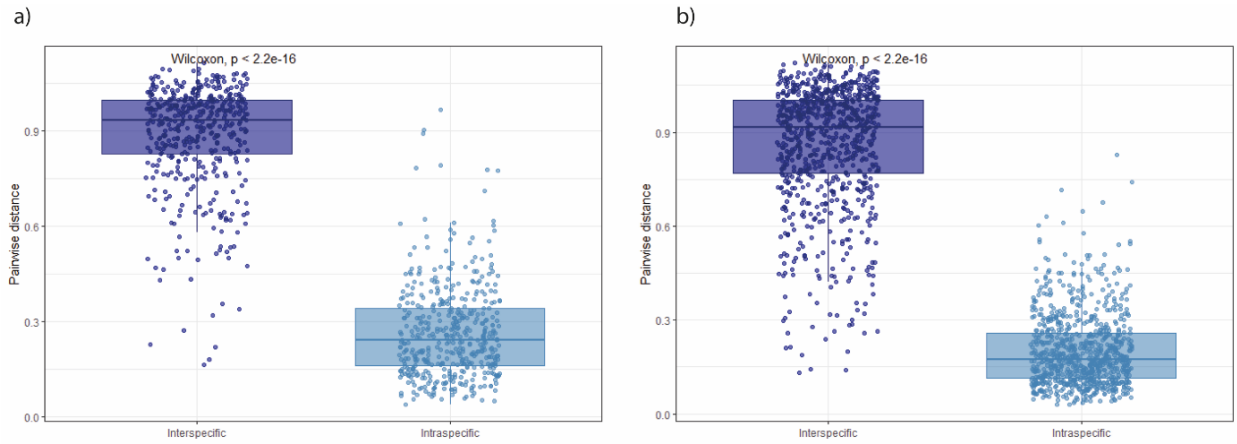

Fig. S2. Pairwise phenotypic Euclidean distances distribution computed from the 20-dimensional morpho space coordinates, for interspecific pairs and intraspecific pairs

separately for a) males and b) females. The same test was repeated for every genus separately, and among the 30 represented genus the interspecific distances were majorly higher than intraspecific distances (male dorsal sides: 23, 2 genera non-significant, 6 genera with only one species, females dorsal side: 17, 3 genus non-significant, 10 genera with only one species).

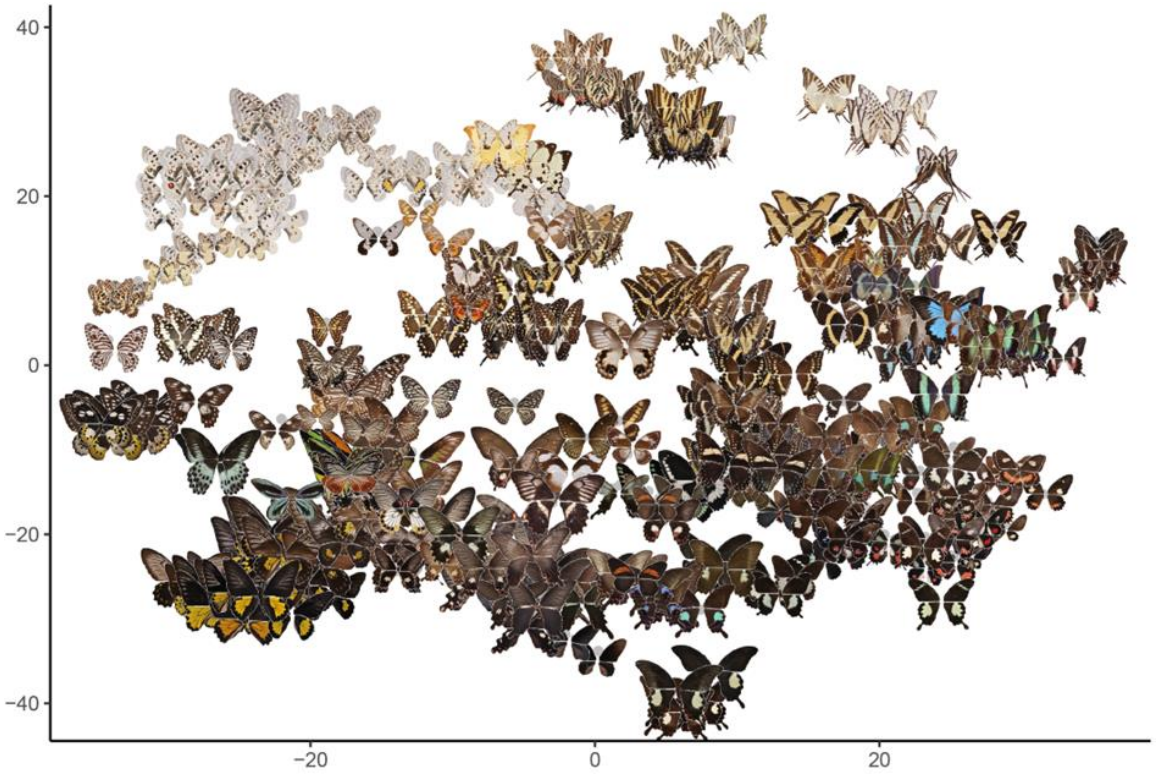

Fig. S3. 2D T-SNE projection of the 2048-dimensional embeddings obtained from the trained SimCLR network. This projection clusters together embeddings that are close in the original space but does not conserve distances, and thus is only helpful to visualize clusters of phenotypes.

2. Impact of shape on phenotypic space. To investigate the impact of wing shape on the final phenotypic space, we generated a dataset constituted by the masks of the wings for each image, filled with random colors. We then trained the method in the exact same way using this dataset, and compared the phenotypic spaces obtained. The goal was to keep the wing shape but randomize color, allowing us to separate the impact of shape from the impact of color. We computed several multivariate correlation measures between the coordinates from the two different trainings using the MatrixCorrelations R package, ranging between -1 and 1. The correlation measures showed little correlation – the maximum value being 0.27 - between the two phenotypic spaces, meaning that the wing shape did not impact much the distance measures.

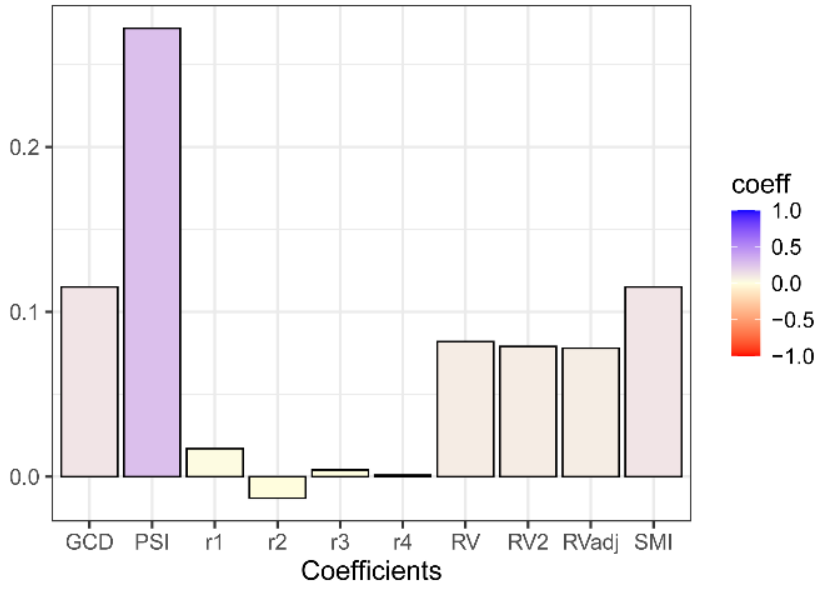

Fig. S4. Various matrix similarity coefficients between the actual phenotypic space and the phenotypic space learned only using wing shape and randomizing colors. PSI : Procruste similarity between the two multidimensional datasets. Other statistics are listed in the MatrixCorrelation package.

3. **Geographical patterns of diversity and phenotypic convergence and divergence.** To test for spatial variation in color pattern disparity, computed as the average squared distance of phenotypes from the centroid of phenotypes, we then gathered the geographical range distribution at the species level for 225 out of the 267 species where both male and female phenotypes were sampled. By mapping the phenotypic disparity on the world map and comparing it with species richness, phylogenetic diversity (Fig. S5), we observed that the areas of highest disparity were in Europe and the Middle East. The spatial variation in disparity therefore did not match the documented hotspots of species richness: the correlation between the level of spatial pattern of disparity and the local phylogenetic diversity was indeed poor (Spearman rank correlation on males: 0.48, females: 0.49). Phylogenetic and morphological diversity indices were computed using the EcoPhyloMapper package in R.

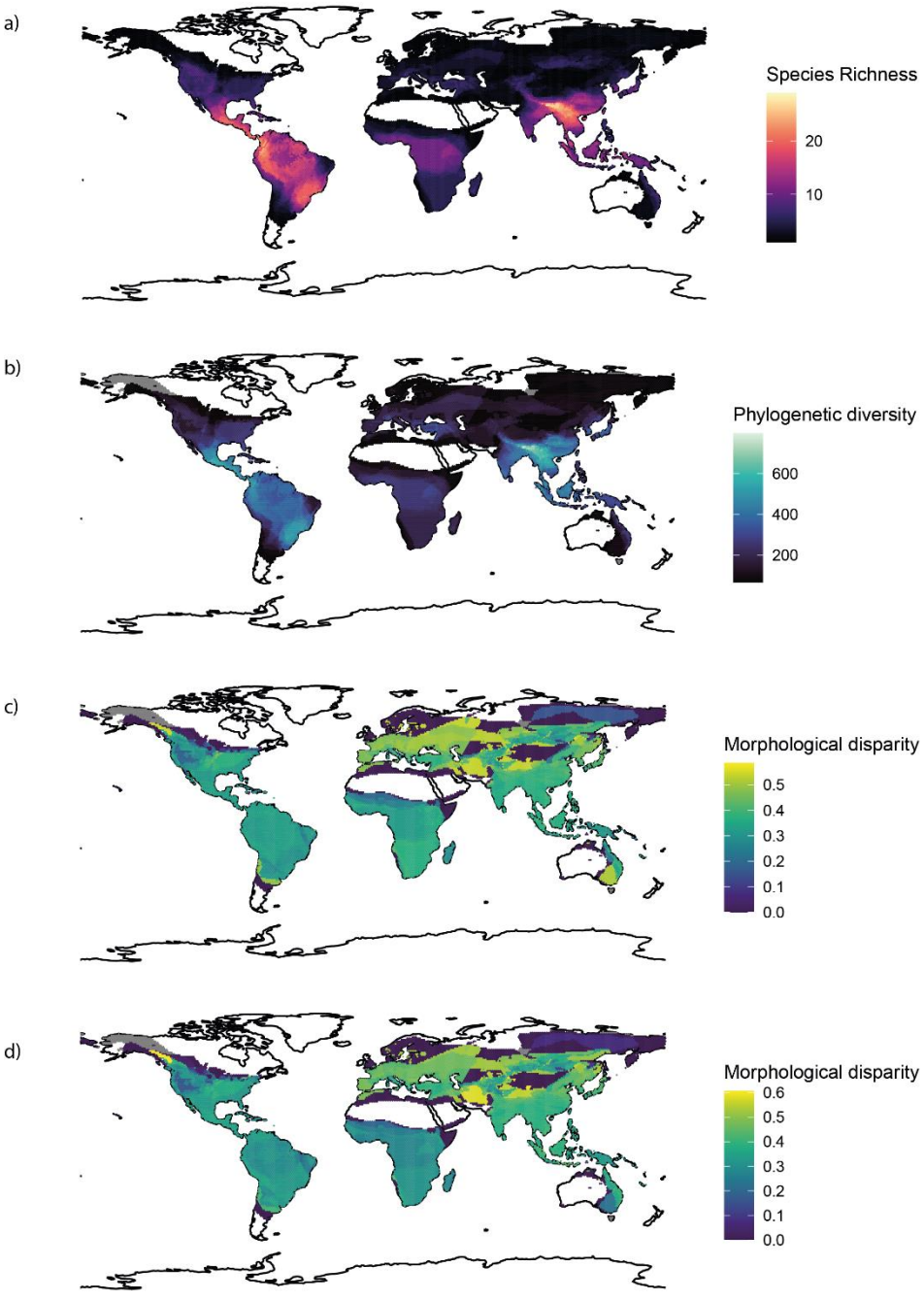

Fig. S5: Geographical mapping of species richness (a), phylogenetic diversity (b), male phenotypic disparity (c), and female phenotypic disparity (d).

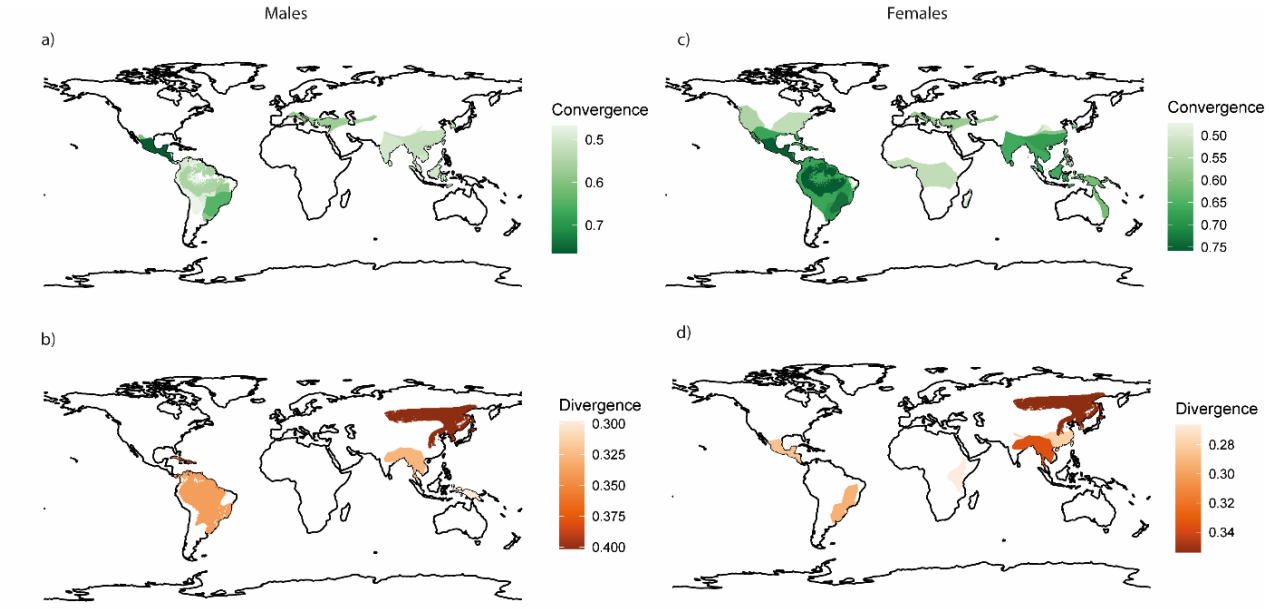

Fig. S6: Geographical mapping of the detected convergence and divergence. Convergent and divergent areas are the intersection between range areas of the pair detected convergent or divergent. The color bar represents the maximum convergence and divergence strength. a) Mapping of male maximum convergence strength b) Mapping of male maximum divergence strength c) Mapping of female maximum convergence strength d) Mapping of female maximum divergence strength

4.Simulation based approach using the mvMorph package. Using the mvMorph package, we fit multivariate models of evolution to our 20 dimensional trait data provided by our machine learning algorithm. We fit a brownian motion model, a Ornstein-Uhlenbeck model and an Early Burst model and evaluated the best fitting model using GIC value. We also fitted a lambda model in order to test for a model that accounts for the significant phylogenetic signal detected in our data. For females, the brownian motion model and the lambda model were the best fitting one (lambda-BM : delta GIC = 0.002,

OU-BM : delta GIC = 2.53, EB-BM : delta GIC = 2.002). For males, the Ornstein-Uhlenbeck model was the best fitting one but very close to the BM model (lambda-OU : delta GIC = 1.872, BM-OU : delta GIC = 1.878, EB-OU = 3.879).

As the OU and BM/lambda models were very close in terms of GIC (delta GIC<2), and the OU model is not justified in our case as it is very unlikely that there exists a single optimum phenotype for the whole Papilionidae family at the global scale, as well as the impossibility to estimate OU parameters in such high dimensions, we selected the lambda model to simulate our traits so as to still account for phylogenetic signal.

We thus simulated trait evolution with the model estimated parameters on the lambda transformed tree for 10000 simulations.

We then compared the observed pairwise mean phenotypic distance for sympatric species standardized by the overall mean pairwise phenotypic distance across all species, with the same statistic in our simulations.

When assuming that sympatry implies at least 20 % of overlap in geographic distribution between the two species (as in the analyses reported in our initial manuscript), we found similar results: significatively smaller phenotypic distances were observed in sympatry compared to simulations for females ($p<0.01$), but for males the signal was less strong and not significant at the 5% level but close ($p=0.06$).

When changing the arbitrary cut-off to assign species pairs to the sympatric and allopatric group, we observed the same trend. Assuming increased levels of overlap to assign species to the sympatric group, makes the test more significant for both males and females, in line with the results obtained on the degree of overlap in the PLMM analyses.

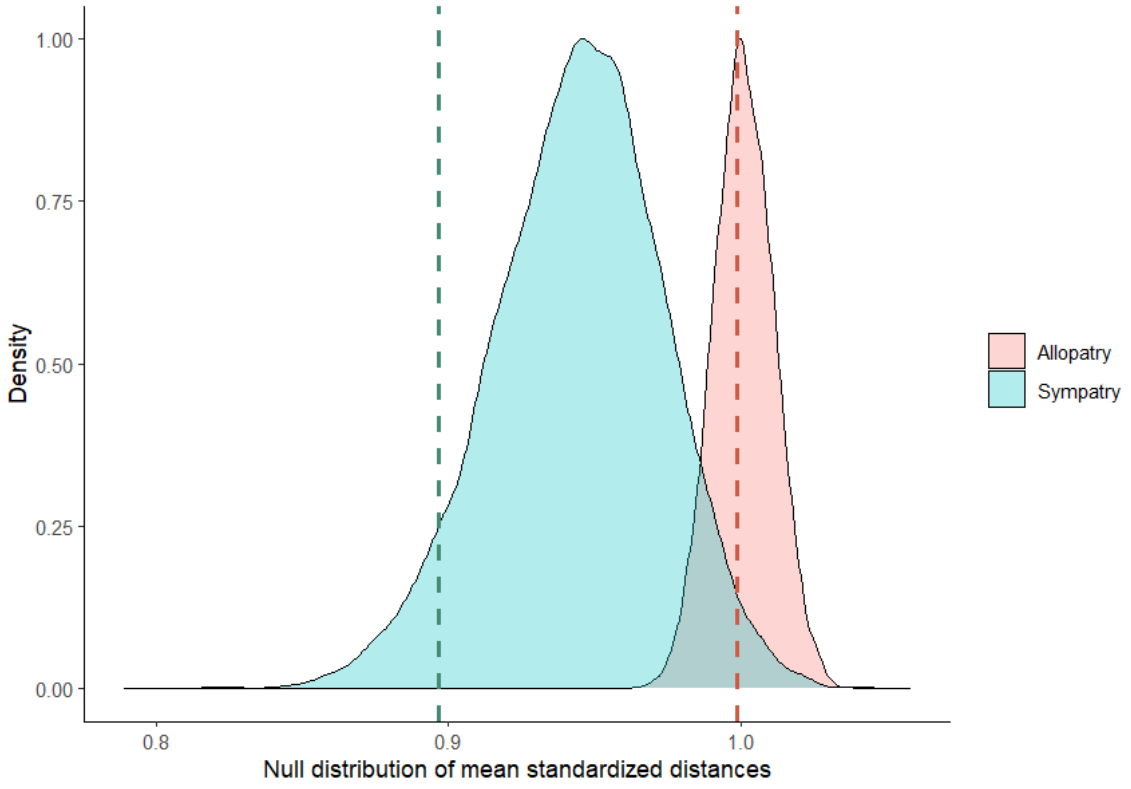

Fig. S7: Distribution of null model simulated mean standardized distances for males after 10,000 simulations for sympatric pairs in blue and allopatric pairs in pink. The dashed blue vertical line corresponds to the observed mean standardized distance of pairs of sympatric species, and the dashed pink vertical line corresponds to the observed mean standardized distance of pairs of allopatric species. (for a sympatry cutoff of 20% : $p=0.06$, 30% : $p=0.07$, 40% : $p<0.05$, 50% : $p<0.05$)

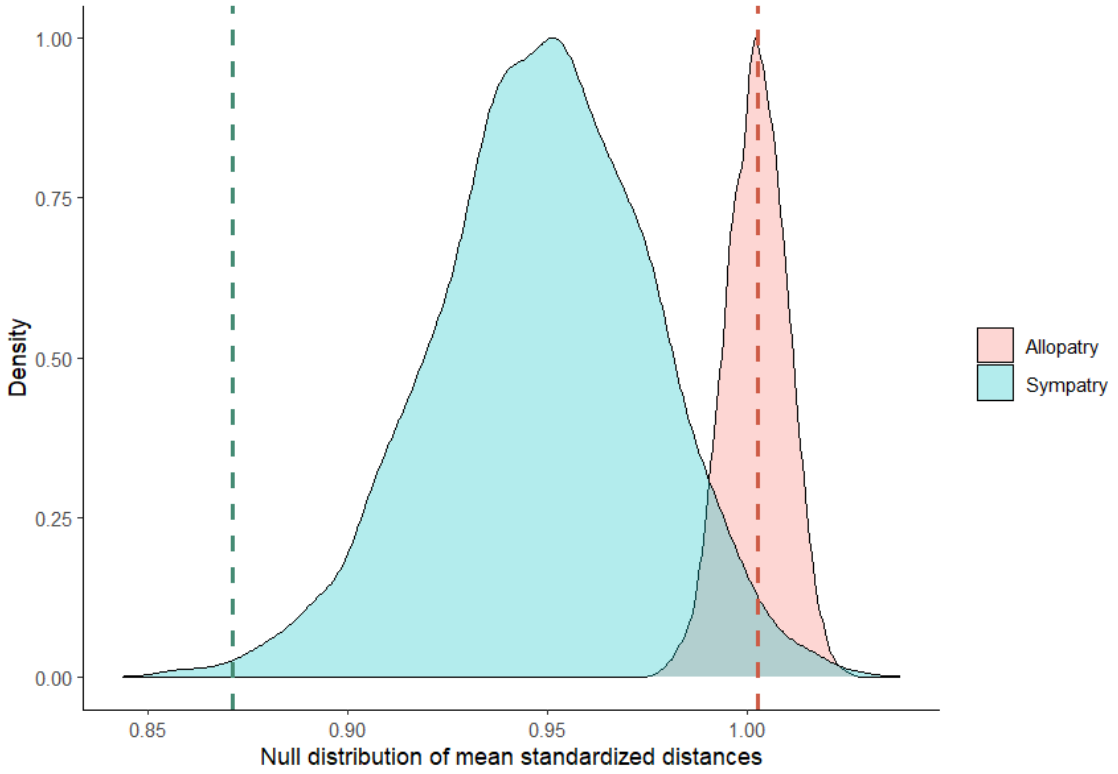

Fig. S8: Distribution of null model simulated mean standardized distances for females after 10,000 simulations for sympatric pairs in blue and allopatric pairs in pink. The dashed blue vertical line corresponds to the observed mean standardized distance of pairs of sympatric species, and the dashed pink vertical line corresponds to the observed mean standardized distance of pairs of allopatric species. (for a sympatry cutoff of 20% : $p<0.01$, 30% : $p<0.01$, 40% : $p<0.001$, 50% : $p<0.01$)

5. Phylogenetic Linear Mixed Models (PLMM) results

We fitted the following model for males and females separately (random effects are in brackets):

Phenotypic distance ~ Overlap + Phylogenetic distance + Phylogenetic distance^2 + Overlap*Phylogenetic distance + Overlap*Phylogenetic distance^2 + (Species 1) + (Species 2)

Where for each data point (pair of species) a link is made to the phylogeny to account for phylogenetic relatedness. Predictor variable were standardized prior to fitting.

The quadratic term of phylogenetic distance was added to account for non-linear accumulation of phenotypic distance regarding phylogenetic distance, following Tobias et. al. (2014 - Nature). A negative term indicates a deceleration of the increase of phenotypic distance with phylogenetic distance (hump-shaped curve).

Variable	Coefficient	p-value
Overlap	-0.01156030	< 2.2e-16
Phylogenetic distance	0.00478737	0.1895
Range overlap:Phylogenetic distance	0.00920400	1.559e-09
Phylogenetic distance²	-0.04647549	< 2.2e-16
Range overlap:Phylogenetic distance²	0.00481016	1.360e-11
Random effect	Variance	p-value (LRT test)
Lineage 1	0.0010229	< 2.2e-16
Lineage 2	0.0019745	< 2.2e-16

Table S3: PLMM results for males

Variable	Coefficient	p-value
Overlap	-0.01541798	< 2.2e-16
Phylogenetic distance	0.00574119	0.1167
Range overlap:Phylogenetic distance	0.00899770	4.56e-09
Phylogenetic distance²	-0.03948247	< 2.2e-16
Range overlap:Phylogenetic distance²	0.00372303	2.02e-07
Random effect	Variance	p-value (LRT test)
Lineage 1	0.0008321	< 2.2e-16
Lineage 2	0.0014118	< 2.2e-16

Table S4: PLMM results for females.

6. Permutations of phenotypes among sex

Following van der Bilj et. al. (*Evolution Letters*, 2020), we permuted the phenotypes between sexes for each species and obtained an expected distribution of the median of our divergence index (*r*) under the assumption of no sex-specific signal (Fig. S9). The red vertical line shows the observed median of *r* in the pairs of sister species of our dataset.

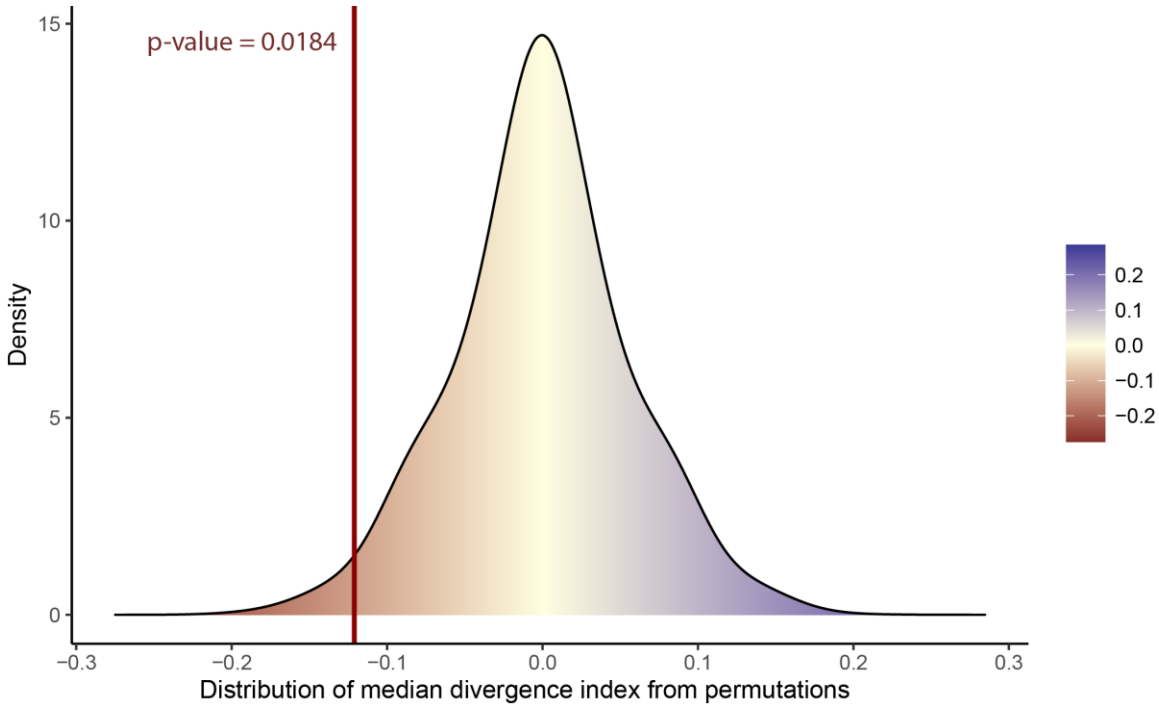

Fig. S9: Distribution of expected median of the divergence index from 10,000 permutations.

7. The impact of phylogenetic correction on detection of convergence and divergence events

As we detected significant phylogenetic signals on wing color variations, we used a linear regression to account for the effect of phylogenetic distances in our analyses. We found that convergent pairs were more distantly related compared to divergent pairs, and that convergence strength was higher than divergence strength. Nevertheless, this phylogenetic correction might bias our detection of convergence and divergence events. When closely-related species display a strong phenotypic similarity, it is challenging to disentangle the

effect of shared ancestry from selection promoting color pattern convergence. Conversely, when distantly-related species display very different phenotypes, the relative effects of phylogenetic distance from divergence due to selection are confounding. The necessary phylogenetic correction therefore limits the detection of (1) convergence events among closely-related species, and (2) divergence among distantly-related species. The greater number of convergence events detected in our study among distantly- vs. closely-related species might stem from decreased reproductive interferences in phylogenetically distant mimetic species but might also stem from the bias induced by the phylogenetic correction. Nevertheless, the significantly greater strength of convergence as compared to divergence in sympatry estimated from our analyses despite phylogenetic correction suggests that natural selection promoting convergence in sympatric species is a significant evolutionary force interacting with neutral divergence. Moreover, the PLMM analysis show similar results regarding phylogenetic distances. Other measurements of convergence, such as the C1 statistic (36), are also a distance based statistic with permutations to assess significance have the same phylogenetic bias. This C1 statistics was recently used to assess convergence among models and mimics within communities (Basu et. al. 2022, *Proceedings of the National Academy of Sciences*) but is more computationally intensive than our method. When computing the C1 statistic for our convergent pairs and comparing our convergent strength, we found similar results.