

Electronic supplementary material of the manuscript entitled ‘Selection on male sex pheromone composition contributes to butterfly reproductive isolation’

Authors: P. M. B. Bacquet^{1*}, O. Brattström^{2*}, H.-L. Wang³, C. E. Allen⁴, C. Löfstedt³, P. M. Brakefield², C. M. Nieberding^{1‡}

Affiliations:

¹ Evolutionary Ecology and Genetics group, Biodiversity Research Centre, Earth and Life Institute, Université catholique de Louvain, Croix du Sud 4-5, 1348 Louvain-la-Neuve, Belgium.

² University Museum of Zoology, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, United Kingdom.

³ Department of Biology, Pheromone group, Lund University, SE-223 62 Lund, Sweden.

⁴ Division of Biological Sciences, University of Montana, Missoula MT 59812, USA.

* These authors contributed equally to this work.

‡ To whom correspondence should be addressed. Phone : +32 (0)10.47.34.88.; fax : +32

(0)10.47.34.90.; e-mail : caroline.nieberding@uclouvain.be

Detailed table of contents of the electronic supplementary material:

Note S1: Sampling and phenotypic scoring

- 1.1. Species Sampling
- 1.2. Scoring pMSP components
- 1.3. Scoring androconia
- 1.4. Scoring eyespots

Note S2: Phylogenetic reconstructions and character mapping

Note S3: Estimating rate of evolution for chemical and morphological traits

- 3.1. One or two rates of evolution for gains and losses
- 3.2. Comparing rates between traits

Note S4: Within- and among-species variability in chemical profiles

- 4.1. Variability of chemical profiles between individuals of the same species
- 4.2. Variability of pMSP composition between *Bicyclus* species

Note S5: Test for Reproductive Character Displacement

- 5.1. Test for RCD in the number of shared traits across species
- 5.2. 'Young' versus 'old' species pairs
- 5.3. Robustness of RCD to potential outliers
- 5.4. Effect of habitat on RCD

Note S6: Test for the presence of a phylogenetic signal for chemical and morphological traits

Note S7: Diversification of pMSPs and androconia is uncoupled

Note S8: Effect of the environment on the diversity of pMSP composition

Note S9: Sympatry and age of the pairs of species - the differential fusion hypothesis

Authors' contributions

Supporting references

Note S1: Sampling and phenotypic scoring

1.1. Species Sampling

We trapped *Bicyclus* butterflies at a total of six sites located in four African countries (ESM figure S1). At each site, we trapped butterflies during a two-week period, and all but five species were sampled from a single site. For each site, we used 10-20 traps and placed them at least 30 m apart. We checked traps daily and moved them regularly around the site, to maximise the captures. At some sites, we placed groups of traps at two locations, never more than 2km apart. At the scale we sampled, we did not observe any differences in species composition across traps for the species included in our study. Although our two-week field sampling scheme might bias our chemical analysis, since the relative amounts of some MSP components can change with age [26], our sampling reflects the actual MSP diversity encountered by interacting individuals in the field.

Our species sampling does not include the Malawian stock population of *B. anynana* from which the selection procedure was designed, because it is not living in sympatry with any of the species sampled for this study and it was reared in the lab for more than thirty years [61]. The *B. anynana* population used in this study was caught in Uganda and is part of a different subspecies, *B. anynana centralis* [62], which does not produce MSP1 and MSP3 found in *B. anynana anynana* from Malawi [25].

1.2. Scoring pMSP components

We conducted chemical analyses using three males and two females per species. We removed both the forewing and hindwing from one side (right or left) of each freshly killed butterfly, (males only) dissected each androconium from each of the wings, and extracted each

androconium in a separate 1.5ml screw-cap vial containing 100 μ l redistilled n-heptane with 1ng/ μ l of (Z)-8-tridecenyl acetate as an internal standard. We placed the remaining tissue of each dissected wing (and wings of females) in a vial of 300 μ l n-heptane with 0.33ng/ μ l of the internal standard. We stored the body in 100% ethanol for genetic analysis and the undissected wings in glassine envelopes for analysis of morphological traits.

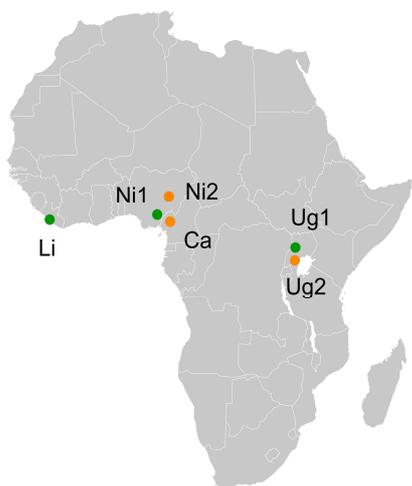


Figure S1: Africa map with the localisation of the sampling sites. The site abbreviations correspond to the tree in figure 2 (main text): Li for Liberia (Sapo); Ni for Nigeria (1: Afi mountains and 2: Yankari); Ca for Cameroon (Fossong Ellelem); Ug for Uganda (1: Kibale and 2: Mburo). Green and orange dots represent tropical forest and savannah sites, respectively.

We analyzed wing extracts on an Agilent 5975 mass-selective detector coupled to an Agilent 6890 gas chromatograph, equipped with a HP-5MS capillary column (30 m x 0.25 mm i.d., and 0.25 μm film thickness; J&W Scientific, USA). The oven temperature was programmed from 80°C for 3 min, then to 210°C at 10°C/min, hold for 12 min and finally to 270°C at 10°C/min, hold for 5 min. Inlet and transfer line temperatures were 250°C and 280°C, respectively, and helium was used as the carrier gas. We only analyzed those compounds with a retention time under 31.8 minutes (retention time of pentacosane using our method); we expect that those compounds with retention time > 31.8 minutes are gustatory chemicals with low volatility. We identified chemical compounds by comparing gas chromatography retention times and mass spectra with authentic standards acquired commercially or prepared by synthesis on both non-polar (HP-5MS) and polar (INNOWax, 30 m x 0.25 mm i.d., and 0.25 μm film thickness; J&W Scientific, USA) columns. The position of double bonds was localised by the DMDS microreaction [63]. Tentative structures were assigned for compounds not fully identified (ESM table S1 in a separate excel file). All 873 GC-MS files are available on Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.768sd>.

If different species tend to share the same sets of structurally related pMSP components, pseudoreplication could occur in our data. To assess the risk for pseudoreplication, we tested whether structurally related components in the pMSP tended to be shared across species, using Fisher's exact tests for all possible pairs of such structurally-related components. Each contingency table contained the number of species for which: both compound *a* and structurally-related compound *b* are present, only compound *a* is present, only compound *b* is present, or, both compound *a* and compound *b* are absent. Out of 231 possible pairs of shared, structurally related components, the test was marginally significant for only one pair ($n = 32$, $or = 12.86$, $p = 0.53$). In this case, the closely related species *B. mollitia* and *B. sylvicolus* both possess the same pair of structurally related components (figure 1). This is probably caused by phylogenetic inertia, and not biosynthetic constraints, since these two compounds are each present alone (but not selected as pMSP components) in two other species, *B. ignobilis* and *B. graueri* (figure 2). For all other possible pairs, the presence of shared, structurally related components in one species was not significantly correlated with the presence of this pair of compounds in other species, eliminating the potential for pseudoreplication in our dataset.

Table S1 (separate excel file): **List of the chemical compounds detected for our sample of 32 *Bicyclus* species and the final selection of 75 pMSP components.**

This table lists all compounds that were present above 10 ng in samples of homologous wing parts in at least two out of three males of the same species. The 75 compounds selected as both male-specific and abundant (pMSP) are highlighted in Column 1. We investigated the chemical structure of all pMSP components; the table details the chemical analyses performed for each compound. Compounds were generally 'fully identified' only in one species (using reference

compounds and DMDS reactions) and then recognised in the other species based on similarity of spectra and retention times. The characteristics of the mass spectra of the few remaining unidentified pMSP components are given. Homologous compounds (based on identical mass spectrum and retention time) to the above listed ones that were observed in additional species were also added to the table. All these homologous compounds were assigned the same reference number to facilitate comparison between species ('compound number' Column in the table). Some compounds identified by Wang et al [47] are not included here because either a) the compound fell out of the window of retention time considered here as a criterion for inclusion as a pMSP component, b) the compound was not male-specific, or c) the compound was not among the most abundant or was not a repeatable compound in the species where it occurred.

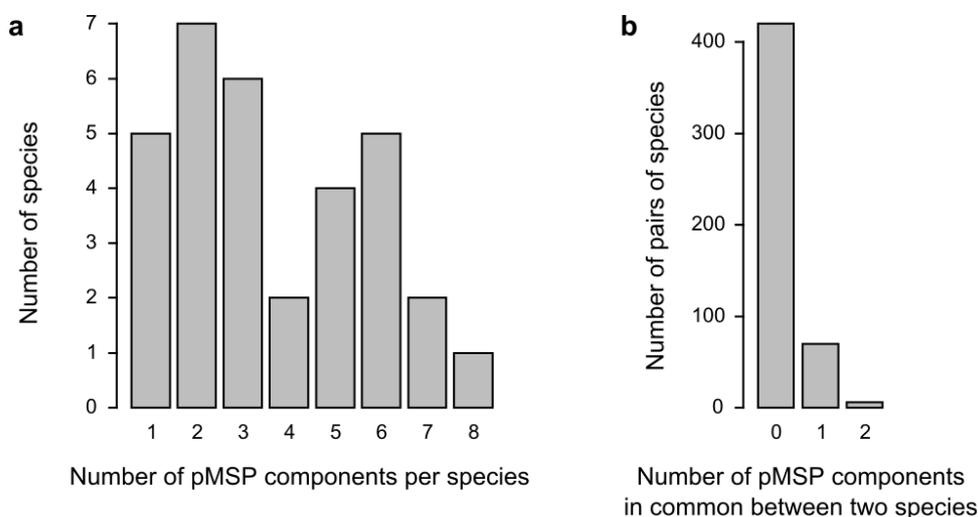


Figure S2: Number of pMSP components per species (a) and number of pMSP components shared per pair of species (b).

1.3. Scoring androconia

Androconia are wing structures consisting of modified scales arranged as ‘patches’ or ‘brushes’. We identified androconia by their location in relation to the wing veins on a given wing surface (ESM figure S3). In total, we identified 20 different structures. Distribution of the androconia on wing surfaces is similar across most *Bicyclus* species; we verified the patterns in our sample by crosschecking 90 currently recognised *Bicyclus* species in museum collections and determined that distribution patterns are robust and not biased by the selection of species included in the pheromone sampling. The ventral hindwing surface lacks androconia in all *Bicyclus* species known. On the dorsal hindwing, androconia lie close to vein intersections; in species with atypical wing venation (e.g., *B. buea*, *B. maritus*), the position of the androconia corresponds to changes in vein position. On the ventral forewing, some androconia are located posterior to the most posterior vein (Vein 1 in Fig. S3), with no interconnecting veins. We were able to reliably identify these posterior forewing androconia and determine their homology across species by taking advantage of the fact that each of these forewing androconia overlaps one of the hindwing androconia when the wings are held in a resting position. Also on the ventral forewing, the androconium located anterior to Vein 1 and the androconium crossing Vein 1 were always scored as separate characters.

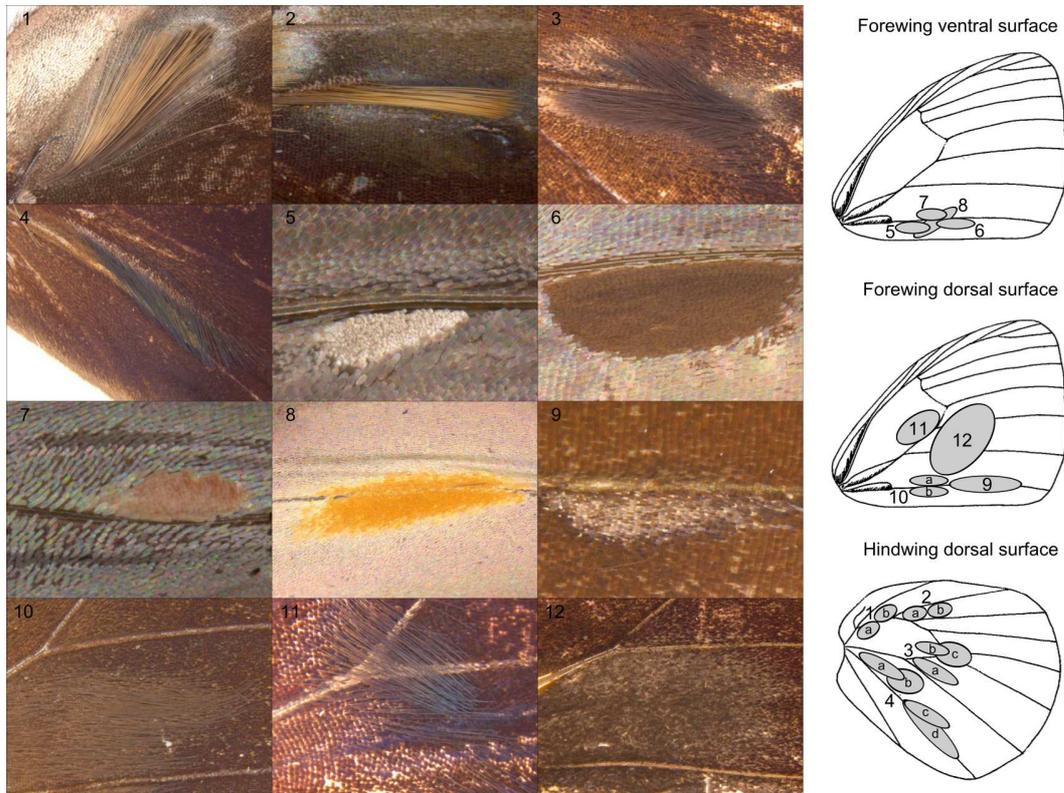


Figure S3: Example picture of each androconial unit and scheme of androconia positions.

Numbers correspond to ESM figure S12. Vein and space numbering follows the simplified ‘English’ numerical system [64]. Dorsal hindwing: (1a) cell brush and (1b) patch in space 7; (2a) brush and (2b) patch in space 6; (3a, 3b) brushes and (3c) patch located around distal part of wing cell; (4a, 4c) brushes and (4b, 4d) patches along space 1c and vein 1b. Ventral forewing: (5) patch under vein 1 covering androconia type 1; (6) patch under vein 1 covering androconia type 2; (7) patch above vein 1; (8) patch stretching across vein 1. Dorsal forewing: (9) hairs along vein 1; (10a, b) patch along vein 1; (11) brush in wing cell; (12) patch of discal scales. These samples belong to the following species: 1) *B. madetes*; 2) *B. golo*; 3) *B. xeneoides*; 4) *B. ignobilis*; 5) *B. anisops*; 6) *B. safitza*; 7) *B. taenias*; 8) *B. medontias*; 9) *B. dentata*; 10) *B. sambulos*; 11) *B. martius*; and 12) *B. mollitia*.

1.4. Scoring eyespots

Eyespots are located on all wing surfaces in *Bicyclus*. However, eyespots on different wing surfaces have likely evolved under different selection pressures, and behavioural studies in the model species *B. anynana* demonstrate that although dorsal forewing eyespots function in mate choice [29,65,66], ventral hindwing eyespots do not, and instead appear to decrease the efficiency of predator attacks [67]. In addition, many *Bicyclus* species exhibit seasonal polyphenism, which affects the size and presence of eyespots in wet season versus dry season forms. Ventral eyespots appear to be most sensitive to changes in temperature (the cue which determines seasonal morphology). To standardize our dataset, we scored eyespot presence or absence for wet season forms only - this corresponds as well to the season in which most reproduction occurs. In cases where our samples were wet-season forms collected from the same locality as those used by Olivier et al (32; which used wet-season forms only), we coded any missing eyespots and bands as 'present' when they were missing from the dry season form, but present in the wet-season form.

Note S2: Phylogenetic reconstructions and character mapping

Monteiro and Pierce [68] and Oliver et al [29] reconstructed the phylogeny of 58 *Bicyclus* species based on Maximum Likelihood and Bayesian methods, using partial sequences of the mitochondrial genes cytochrome oxidase I and II (COI, COII) and of the nuclear gene elongation factor 1 α (EF-1 α). Our work added two taxa (*B. ephorus* and *B. sylvicolus*) and completed data for a third (*B. sangmelinae*).

Sequencing: Following an adjusted protocol from Monteiro and Pierce [68], we sequenced a part of the mitochondrial gene cytochrome oxidase I (COI) and the nuclear gene elongation factor 1 α (EF-1 α). In *B. sangmelinae*, we failed to obtain the first half of the COI sequence, therefore this part was kept from the dataset of Monteiro and Pierce [68]. The resulting sequences are available on GenBank with accession numbers KC786271 to KC786277.

Model of substitution: We used the AICc and BIC criteria in the program jModeltest 0.1.1 [69,70] to select of the model of sequence evolution to use in our phylogenetic reconstruction. The GTR+G model was found to be the most suitable across three different methods of data partitioning: one partition with the three genes; two partitions to separate mitochondrial genes from the nuclear gene; and three partitions with one separate model for each of the three genes.

Bayesian reconstruction with the different partitions: We first conducted a Bayesian reconstruction of the tree using the parallel version of the program MrBayes 3.1.2 [71–73]. The Metropolis-coupled, Markov chain Monte Carlo (MCMCMC) process included four chains, three heated and one cold with a temperature parameter of 0.15. Starting from random trees, we simultaneously performed two independent runs for 20 million generations each, and sampled the chains every 2000 generations yielding a total of 10,000 samples for each run. The convergence of the two runs was assessed visually with the program Tracer 1.5 [74]; the

standard deviation of the split frequencies stabilized around 0.003 halfway through the run time. Therefore we obtained the 50% majority-rule consensus from the latter 10 million generations (10000 samples in total). Among the three methods of data partitioning, the model with two partitions (nuclear and mitochondrial sequences) received very strong support from the posterior probabilities (Bayes factors of 503 and 30 when compared to the models with one or three partitions, respectively). The Bayes factors are twice the difference of the logarithm of the harmonic mean of the likelihoods of the two models. A value greater than 2 is taken as ‘positive’ evidence that the model is better, greater than 5 as ‘strong’ and greater than 10 as ‘very strong’ evidence.

Maximum likelihood reconstruction and bootstrap: Since posterior probabilities can overestimate the node support [75], nonparametric bootstrapping ($n = 2000$ pseudoreplications) of the Bayesian consensus tree was performed using the online version of the program GARLI (Genetic Algorithm for Rapid Likelihood Inference [76,77]). The Bayesian and Maximum Likelihood methods gave compatible results, although some nodes remain unresolved with the Maximum Likelihood method. As expected, the tree we obtained showed strong similarity to the recent Bayesian reconstruction [29]. For all our analyses, we therefore used the Bayesian trees, pruned to only contain the taxa with available data for the pMSP. Resolution of the tree and bootstrap support of the topology were strong (BT and PP shown in ESM figure S4).

Character mapping: to represent the changes of presence and absence of pMSP components and androconia along the phylogenetic tree, we used the maximum likelihood approach with the package StochChar v1.1 in Mesquite 2.75 [78] with a symmetrical continuous time Markov model of evolution (figure 2).

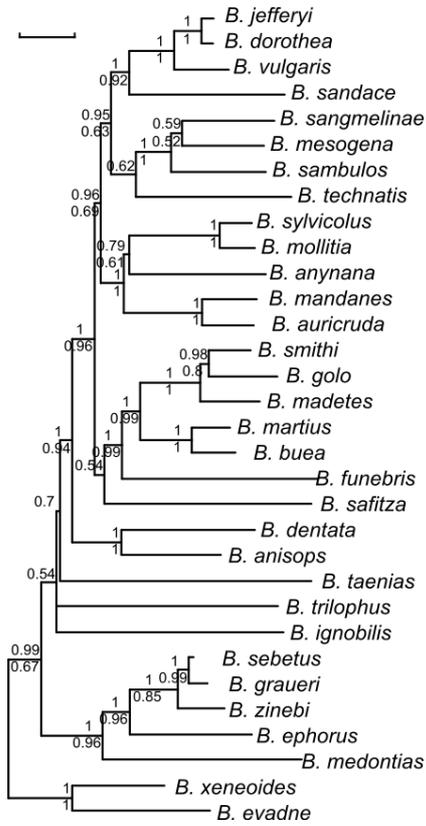


Figure S4: Pruned Bayesian consensus tree.

Numbers represent posterior probabilities (above) and bootstrap values (below) when higher than 50%. The scale bar above the tree represents 1% of genetic divergence.

Note S3: Estimating rate of evolution for chemical and morphological traits

3.1. One versus two rates of evolution for gains and losses

Estimated rates are similar when we estimate gains and losses as a single parameter (Figure 4), or two separate parameters (ESM figure S5). For the ventral eyespots and androconia, the model containing different rates of gains and losses has stronger support than the model containing a single, common rate (Bayes factors of 10.47 and 3.32, respectively; see ESM note S2 for definition of Bayes factor). The rate values indicate that the frequency of gains is higher than the frequency of losses, meaning that both traits have diversified during the radiation of *Bicyclus*. In contrast, for dorsal eyespots and pMSP components, the model with a common value for the rates of trait gains and losses is better supported than with different rates of gains and losses (Bayes factors of 2.40 and 2.18, respectively).

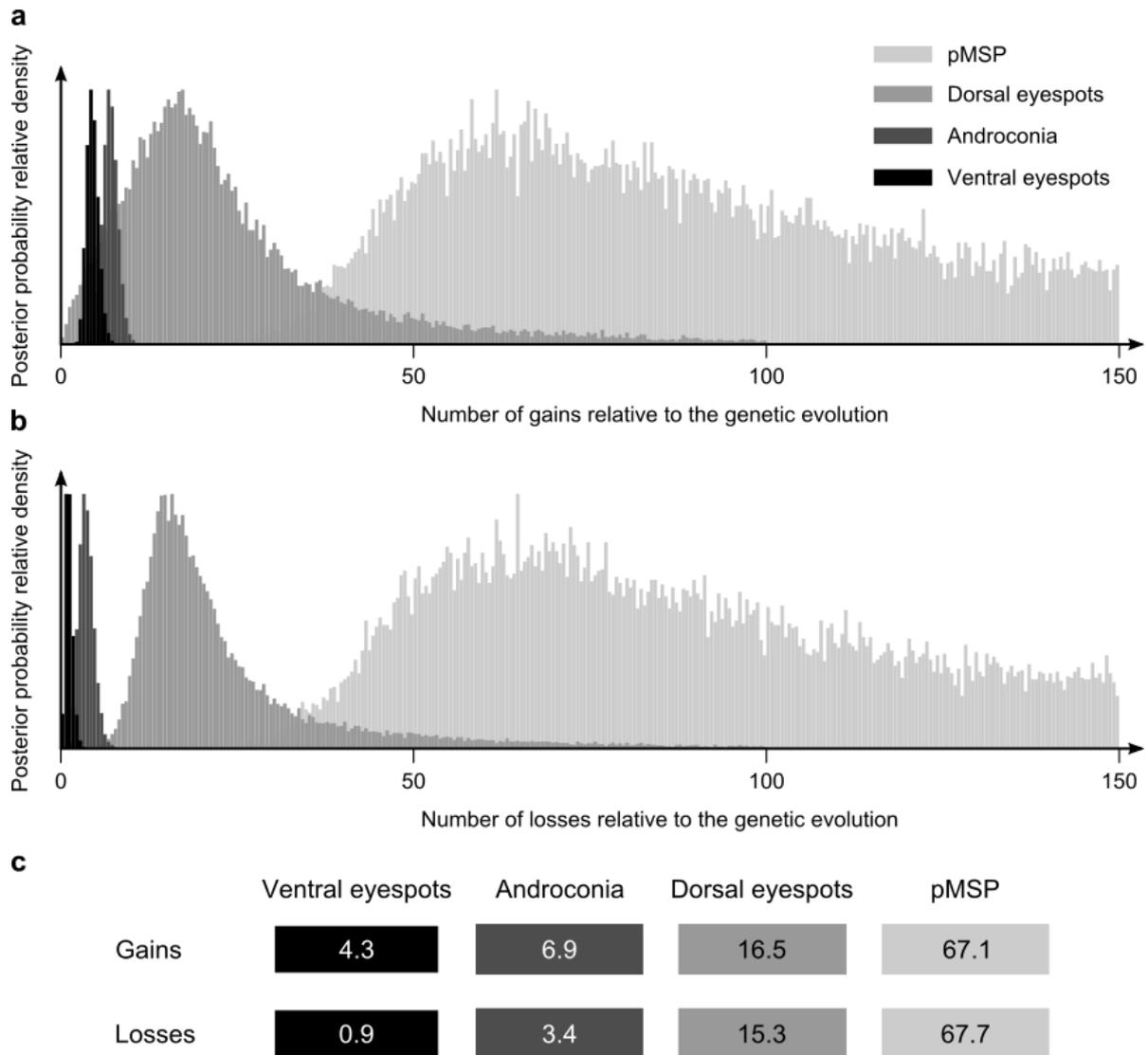


Figure S5: Posterior distribution of evolutionary rates when gains and losses are estimated as two independent parameters.

The traits (from left to right) are: Ventral eyespots, androconia, dorsal eyespots, and pMSP components. The most probable rates of gains and losses (**a** and **b**, respectively; number of change relative to genetic evolution) are presented in panel **c**.

3.2. Comparing rates between traits

Our results show that, across species, pMSP composition turns over four times faster than dorsal eyespot patterns, which in turn occurs twice as fast as androconial characters (figure 4).

However, it is important to interpret these results with caution. The way we coded the morphological traits, distinguishing them only on the basis of their position on the wing (presence, absence), may account in part for their lower rate of evolution compared to pMSP. Indeed, androconia and eyespots can differ in many features in addition to their position on the wing, including shape or UV-reflectance, the latter being a target of sexual selection for *B. anynana* eyespots [65]. Moreover, the way we coded the pMSP diversity may also account for their observed higher rate of evolution: the pMSP components were distinguished on the basis of their unique chemical structure. Yet, the same pMSP chemical precursor can lead to different pMSP components that are considered different traits (here present or absent given our coding) rather than different character states. This increased the number of scored traits compared to morphological traits, potentially inflating the real rate of evolution of pMSP compared to morphological traits. These limitations are inherent to the analysis of rates of evolution of independent traits. We considered the chemical distinctiveness of each pMSP component biologically meaningful since the interaction between sex pheromones and their olfactory receptors are highly specific [20].

Note S4: Within- and among-species variability in chemical profiles

4.1. Between individuals of the same species

We used Spearman-rank correlations to estimate the similarity of whole chemical profiles between different individuals within each species. We estimated the correlation for the amount of each chemical in the profile for a) the three males sampled in each species, and b) the two females sampled in each species, accounting for where each was localized on the wings (androconia, rest of wing). Results therefore include four correlation estimates per species (a total of 128 for the 32 species). We performed this analysis using: 1) the complete list of chemicals; 2) the 50% and 10% most abundant chemicals per species for males only; and, 3) the pMSP components for males only (ESM figure S6). Chemical profiles were found to be repeatable between conspecifics and this similarity increases when the most abundant compounds are considered separately (ESM figure S6). For some species, the similarity was low between conspecifics but we considered this as natural variability (potentially due to age differences between individuals [26]) and kept these species in the data. This is conservative since the confidence on their assigned MSP composition is weaker than for other species and risk to obscure the observed effect of RCD.

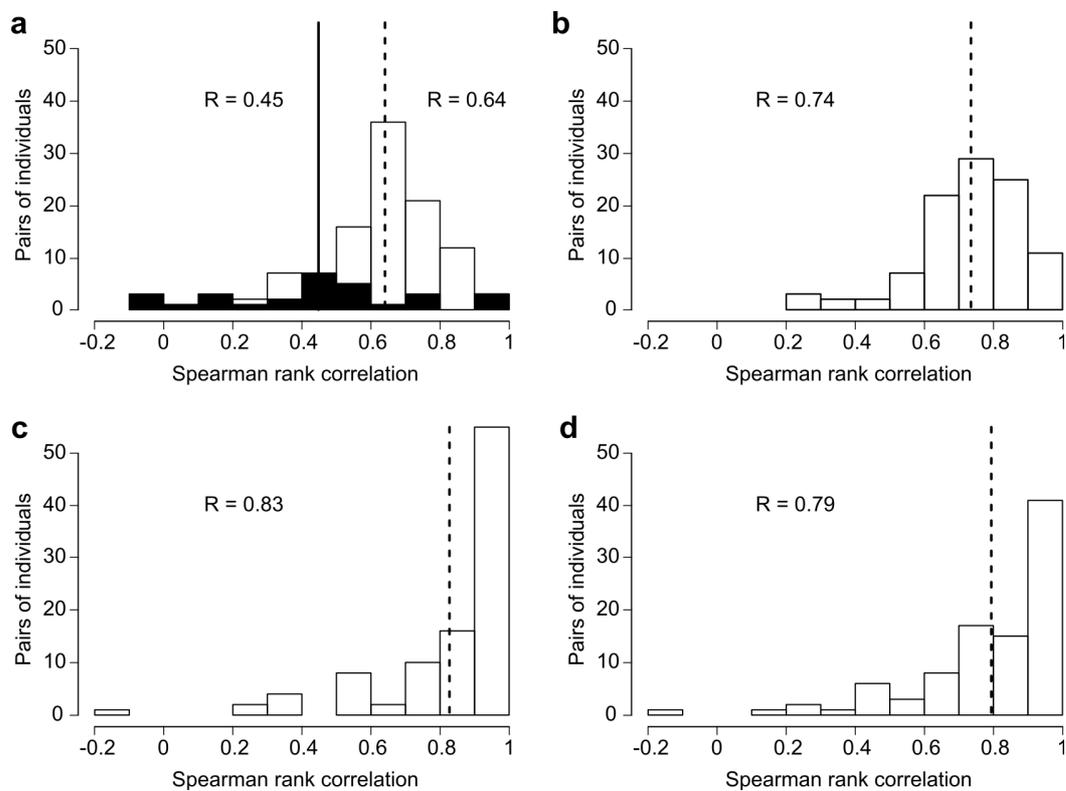


Figure S6: Similarity of chemical profiles among conspecific individuals.

Distribution of the Spearman-rank correlations (R) between individuals of the same sex for different subsets of compounds (mean values of R represented by vertical bars). **a**: all compounds (males and females in white and black, respectively); **b** and **c**: the 50% and 10% most abundant male compounds (males only); **d**: the selection of 75 pMSP components (males only).

4.2. Between individuals of different species

We also used Spearman-rank correlations to compare the similarity of chemical profiles when a) pairs of individuals were drawn from the same species, versus b) pairs of individuals were drawn from different species. For males, the correlation was significantly higher when males were conspecifics (ESM figure S7a), versus when males belonged to different species (ESM figure S7d). For females, chemical profiles were more similar across species compared with males (ESM figures S7c and f). We conducted additional analyses to determine whether this result (more differences in heterospecific profiles among males than among females) was due to the larger number of chemical compounds identified in males (as most pMSP compounds are strictly male-specific). To do this, we resampled a subset of male chemical compounds to equal the average number of chemical compounds typical of female chemical profiles and showed that the gap between the distributions of conspecific and heterospecific correlations in males persisted (ESM figure S7b and e). Moreover, Wilcoxon rank test is more significant for males ($W = 423006, P < 2.2e-16$ and $W = 421267, P < 2.2e-16$ for all pMSP components and for the resampled data, respectively) than for females ($W = 28793, P = 2.5e-05$). These results suggest that the large differentiation of male chemical profiles between species could be used in species recognition.

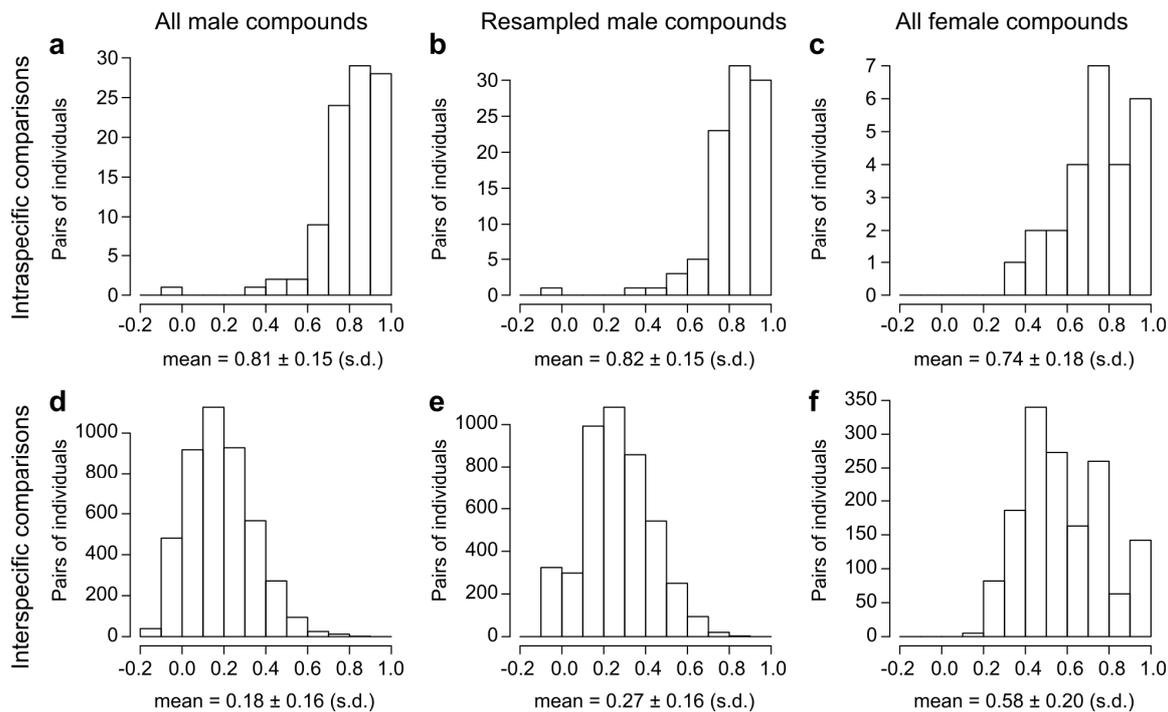


Figure S7: Similarity of chemical profiles among individuals of the same and of different species.

Spearman rank correlations on the amounts of all pMSP components between pairs of conspecific (top row), or heterospecific (bottom row) individuals. Of note, panel **a** is different from ESM figure S6d because in the latter, the localisation of the compounds on the wing is taken into account while here only the total amount per individual is used.

Note S5: Test for Reproductive Character Displacement

5.1. Test for RCD in the number of shared traits across species

The number of shared pMSP components was determined only by phylogenetic distance (permutation test; 9999 permutations; $n = 496$ pairs of species; effect size = -2.49; $P < 0.01$; ESM figure S8). We observed a similar pattern for the androconia (effect size = -14.02; $P < 0.01$) but not for dorsal and ventral eyespots ($P = 0.52$ and 0.60 , respectively; but see ESM note S6).

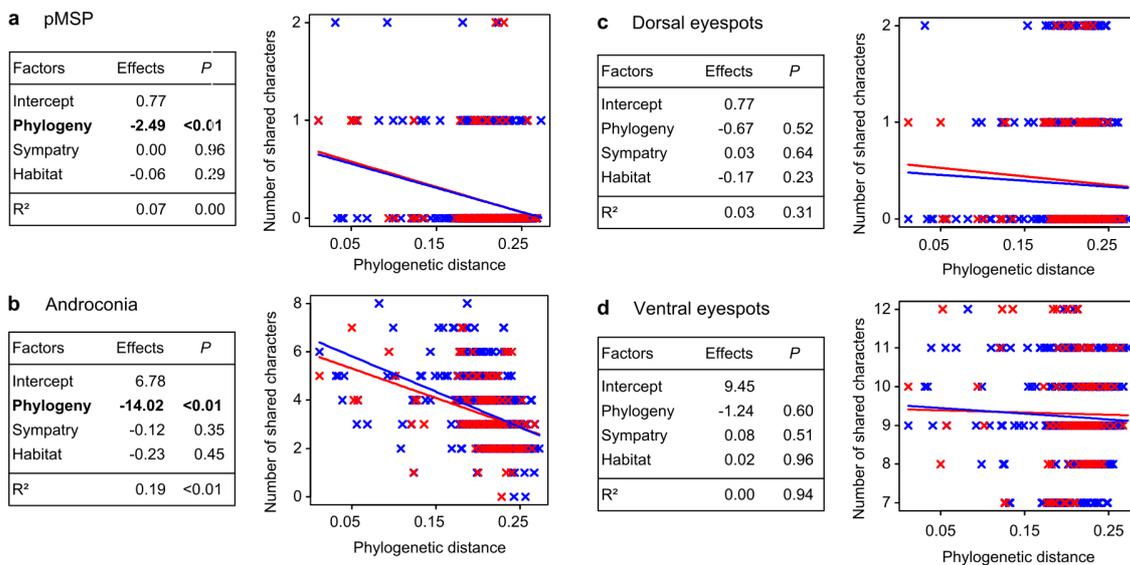


Figure S8: Sympatry, phylogeny and habitat effects on the number of shared characters between pairs of species.

For each trait (**a**: pMSP composition, **b**: androconia, **c**: dorsal eyespots, and **d**: ventral eyespots), statistical output of the following model is given in a table on the left: *number of shared characters per pair of species ~ intercept + phylogeny + sympatry + habitat* (Methods). Significant effects are represented in bold. On the right, graphs show the numbers of shared

characters between pairs of species against their phylogenetic distance, using the following model: *number of shared characters per pair of species* ~ *intercept* + *phylogeny* + *sympatry* + *habitat* + *phylogeny* : *sympatry*. Sympatric and allopatric pairs of species are represented using red and blue crosses, respectively. These graphs thus include the interaction term for illustration sake and the effect of the habitat is averaged.

5.2. 'Young' versus 'old' species pairs

Figure 3a shows that the effect of sympatry on differentiation of pMSP composition is strongest when species pairs are closely related phylogenetically, compared to older species pairs.

Therefore, we recalculated the model coefficients after splitting the data between 'young' and 'old' pairs of species around the phylogenetic distance of 0.15 (figure 3a). Significance of the model parameters was obtained after permutation of the whole dataset, and fitting again the model on the two groups of pairs (ESM table S2). Results show that the effect size of sympatry is much larger when only young pairs of species are included in the analysis (1.77), compared to when either all species pairs (effect size of 0.66, figure 3a) or only older pairs of species (0.53) are considered. Separate permutation of the two partial datasets gave equivalent results (results not shown).

Table S2: Effect of the phylogenetic distance, the sympatry, and the habitat type on the number of pMSP differences for the young and old pairs of species.

	Young pairs ($n = 38$)		Old pairs ($n = 458$)	
Factors	Effects	<i>P</i>	Effects	<i>P</i>
Intercept	2.50		5.62	
Phylogeny	5.48	0.59	-5.57	0.70
Sympatry	1.77	0.02	0.53	0.04
Habitat	2.03	0.07	1.59	0.06
R ²	0.32	0.22	0.12	0.07

Significant effects are in bold.

5.3. Robustness of RCD to potential outliers

We found a greater number of pMSP component differences across sympatric species compared to allopatric species, and this result is robust even when three potential outlier pairs of species are deleted (ESM figure S9). It is important to note that these three pairs of species are not phylogenetically close - in fact, they represent independent evolutionary events at the scale of the genus and should not be discarded from the analysis. It is interesting to note, however, that the effect of forest habitat on the number of pMSP differences between species is significant with the 3 pairs removed from the data.

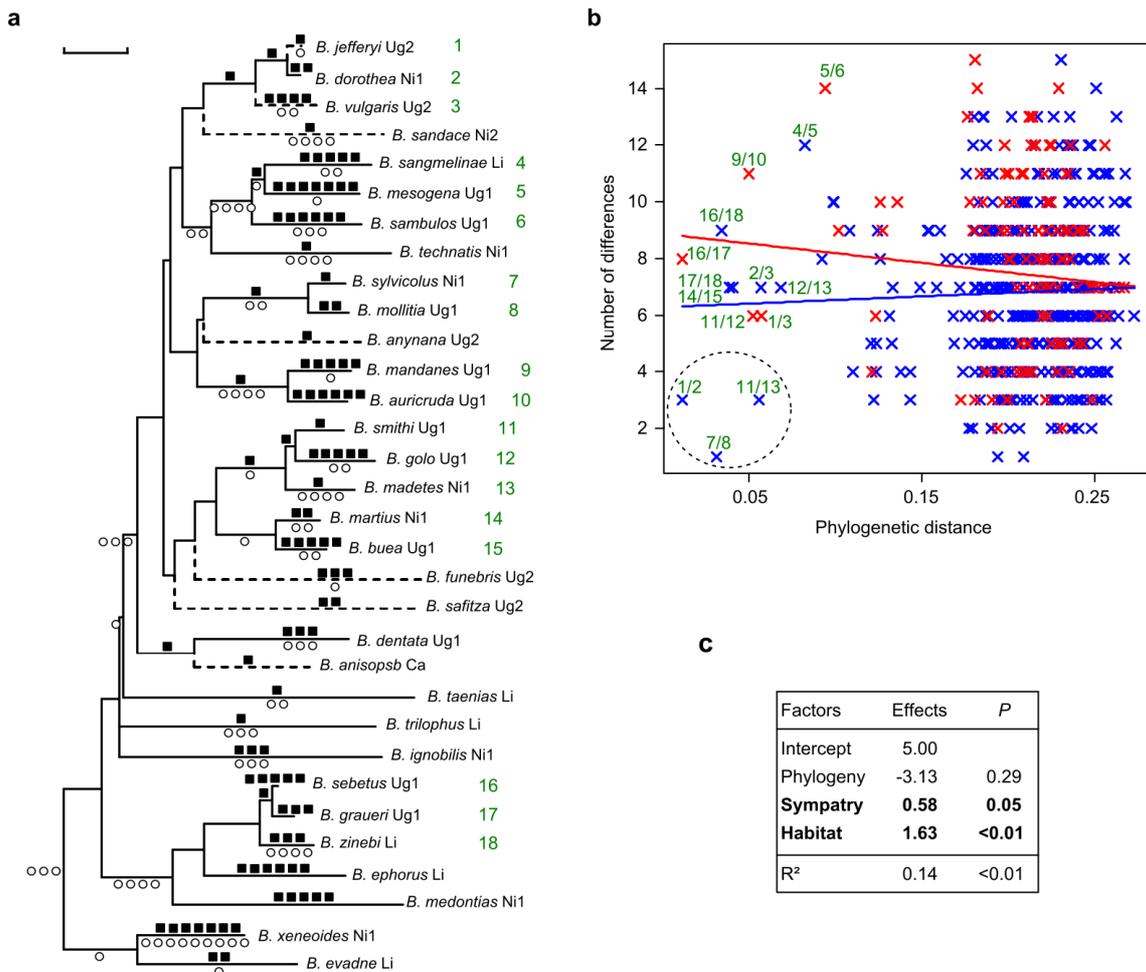


Figure S9: Composition of the youngest pairs of species and robustness of the RCD pattern to the exclusion of potential outlier pairs.

a: Pruned Bayesian consensus tree with inferred changes for pMSP components (filled squares) and androconia (open circles; ESM note S2). The scale bar above the tree represents 1% of genetic divergence. Plain and dotted tree branches represent forest and savannah habitats respectively. **b:** Numbers of different pMSP components between pairs of species against their phylogenetic distance. Sympatric and allopatric pairs of species are represented by red and blue crosses. Lines represent the prediction of the following linear model: *number of differences per pair of species* ~ *intercept* + *phylogeny* + *sympatry* + *habitat* + *phylogeny: sympatry*; for representation sake the effect of the habitat is averaged. Green numbers indicate the position of the youngest pairs of species in the phylogenetic tree. **c:** Results of the corresponding regression on distance matrices without the three potential outlier pairs surrounded in panel **b** by the dotted circle.

5.4. Effect of habitat on RCD

Species inhabiting forest possessed greater numbers of pMSP components and more androconia. While this affects the number of differences we found in these traits (ESM figure S10a,b), it does not alter the effect of sympatry on the number of differences in traits.

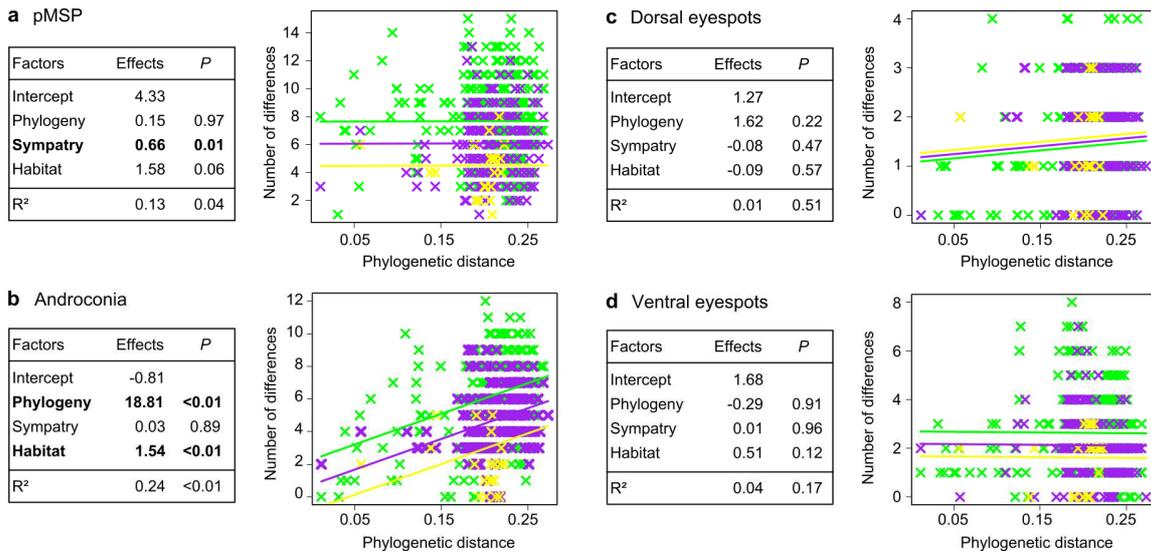


Figure S10: Effect of the habitat type on the number of different characters between species.

For each trait (**a**: pMSP composition, **b**: androconia, **c**: dorsal eyespots, and **d**: ventral eyespots), the statistical output of the following model is given on the left: *number of differences per pair of species* ~ *intercept + phylogeny + sympatry + habitat* (Methods; same model as for figure 3). Significant effects are represented in bold. On the right, graphs show the numbers of trait differences between pairs of species against their phylogenetic distance for the same model. Here the effect of sympatry is averaged and the effect of habitat is represented in yellow, green and purple for species pairs from savannah, from forest or from mixed habitats, respectively.

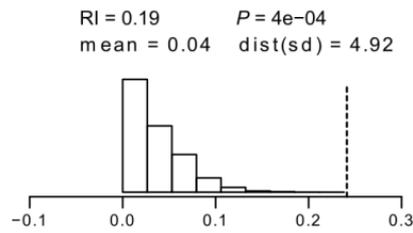
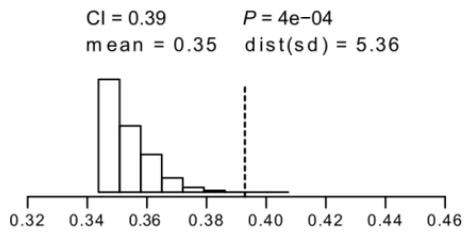
Note S6: Test for the presence of a phylogenetic signal for chemical and morphological traits

We tested whether a phylogenetic signal was present for all chemical and morphological traits using two methods. First, for each trait, we compared the fit of a symmetrical Markov model of evolution on two different phylogenetic trees following Mooers *et al.* [79]. The traits were allowed to evolve either along the branches of the real phylogenetic tree, or along a star-like tree (equivalent to a scenario of no shared history between species). When a phylogenetic signal exists for the trait, the real tree should explain the trait evolution better than the star-like tree. We compared the performance of the two trees using Bayes factors in the program BayesTraits (see the definition of Bayes factor in ESM note S2). For all the traits, the real phylogenetic tree allowed a better fit of the model than the star-like tree (symmetrical Markov model, Bayes factors: 5.36, 100.89, 5.09 and 33.61 for pMSP, androconia, dorsal and ventral eyespots, respectively), showing that more closely related species tend to share more characters than more distantly related ones.

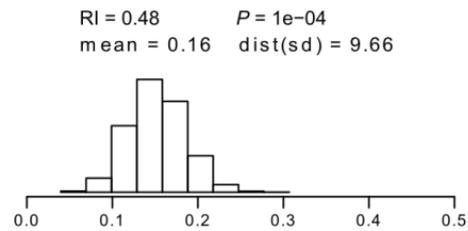
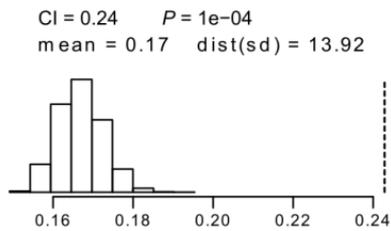
Second, we measured the consistency (CI) [80] and retention indices (RI) [81] which assess the congruence between the data and the tree under the hypothesis of maximum parsimony. This analysis was run with R (version 2.13.1 for windows) [38] and the package Phangorn (version 1.4-1) [82] and Picante (version 1.3-0) [83]. Since the CI is the ratio between the minimum possible number of character changes in the phylogeny and the observed number, it is sensitive to autapomorphies (characters present in only one species) which inflates its value [84]. The number of autapomorphies has no effect on RI. We therefore only considered the shared characters in the analysis. To test for the significance of these measures, we compared them to a null distribution obtained by repeatedly computing the indices after permutation (9999

replicates) of the associations between sequences of characters and species. This method allows testing for the phylogenetic signal without disturbing the potential association of characters in each blend or combination of androconia or eyespots. It thus avoids the production of unrealistic datasets with, for example, some species having no androconia or too many chemical compounds. Because of the difference between the datasets in terms of percentage of shared characters, the expected values of CI and RI under the null hypothesis of no congruence between tree and trait are different (see the mean of the null distributions on ESM figure S11). Since the *P* values were very small for both traits, we measured the distance of the observed value to the null distribution relative to the standard deviation of the null distribution. These indices showed that the congruence between the characters and the phylogeny is stronger in androconia than in pMSP, and stronger in ventral than in dorsal eyespots (no signal in the latter; ESM figure S11). The strength of the phylogenetic signal measured with CI and RI followed the same ranking as the sizes of the Bayes factors obtained with the model comparison.

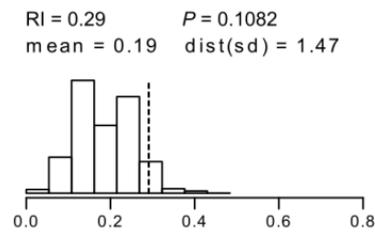
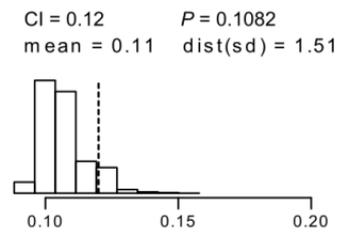
pMSP



androconia



dorsal eyespots



ventral eyespots

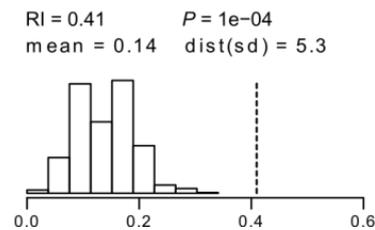
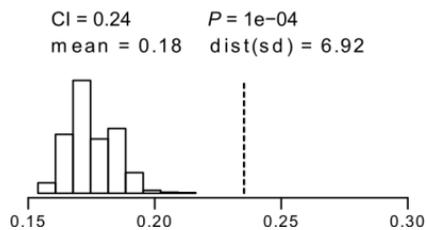


Figure S11: Evidence for the presence of a phylogenetic signal in chemical and morphological traits.

The observed consistency index (CI) and retention index (RI) values (left and right graphs, respectively) are represented by vertical dotted lines, and their null distributions are represented by histograms. The P values, the mean values of the null distributions, and scaled distance

between the observation and the null distribution [dist(sd)] are indicated above each graph. For the convenience of visual comparison between traits, abscissas are aligned and scaled with the standard deviation of each null distribution.

Note S7: Diversification of pMSPs and androconia is uncoupled

Androconia have been used as a proxy to assess the existence and the importance of sex pheromone communication in male Lepidoptera [85], and are a key taxonomic trait in *Bicyclus* [24]. Our analyses show that the number of differences in both androconial traits and pMSP components between pairs of species is affected by habitat in the same way: there are more differences between pairs of forest species than between pairs of savannah species (figure 3). Because habitat affects both traits in the same way, we further investigated whether specific androconia were associated with the presence of specific pMSP components. To conduct this analysis, we identified twelve ‘androconial units’ by classifying as a single unit pairs of brushes which merge together despite originating in more than one wing space, and faint patches lying directly under brushes (see ESM figure S12 for positions of these units). From the abundance data it was clear that most compounds were either found exclusively in a single androconial unit per species, or if found in several only as a trace in all positions except on dominant position. We therefore investigated which androconial unit contained the most of each selected pMSP for each species. The majority of pMSP components (65%) occurred in higher amounts in the ‘cell brush’ unit (unit 1; ESM figure S12). The cell brush occurs in 31 of the 32 species we investigated (absent in *B. buea*), and this structure alone appears to contain a high proportion of the pMSP diversity we identified. The remaining 25% of the pMSP components occurred in parts of the wings which did not contain androconia. In most cases, the remaining androconial units (units 2-12; there are 1-5 units per species, see ESM figure S12) generally lacked significant amounts of any pMSP components. Despite the fact that the number of androconial units seems to have increased during the diversification of the genus (ESM figure S5), we did not find a one-to-one association between pMSP components and specific morphological

structures. Most androconia may have lost a function in olfactory communication, or perhaps never had one. Our results indicate that the current diversity of androconia and pMSP components are independent and that these elaborate and fixed morphological structures may have a role in another, as yet unknown, function maintained by selection.

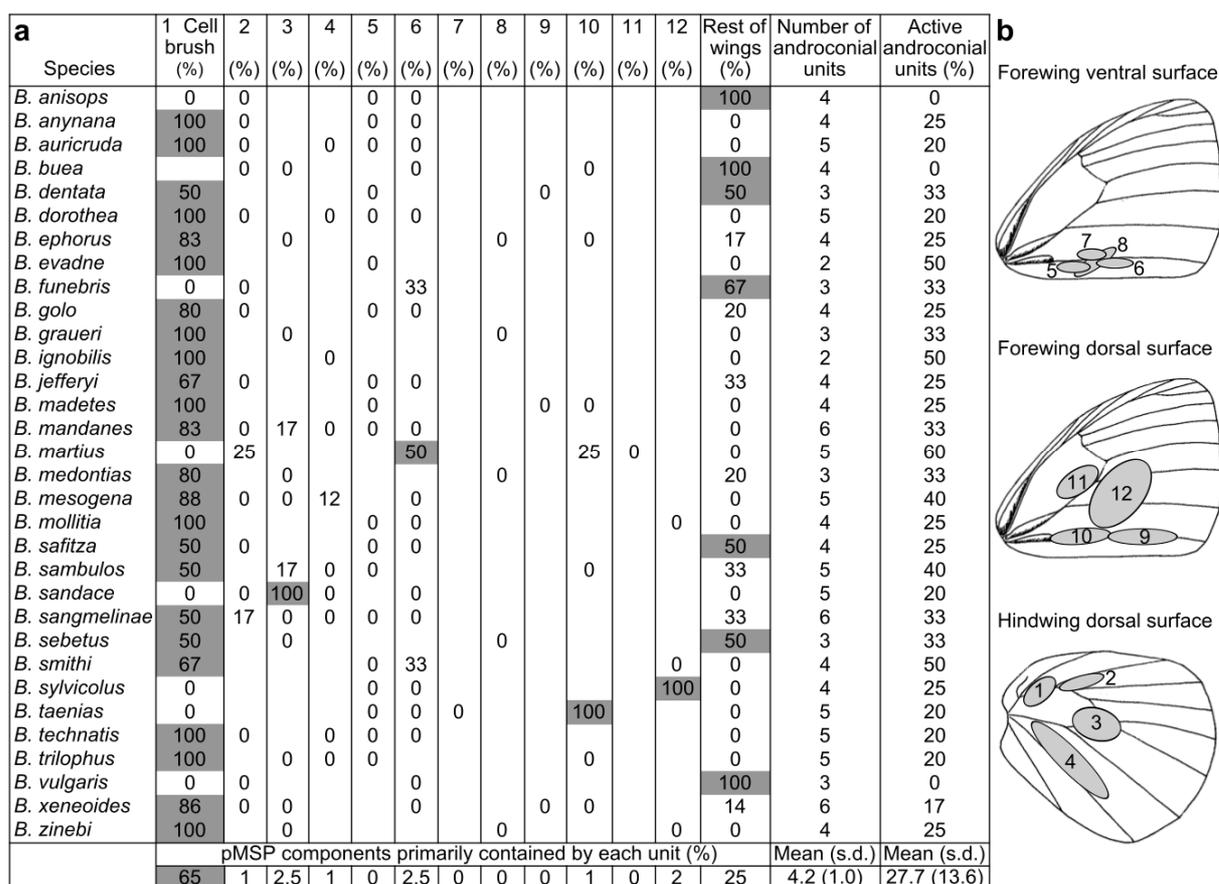


Figure S12: Distribution of pMSP components among androconial units on *Bicyclus* wings.

a: Columns represent each of the 12 androconial units (association of androconial hairpencil and patch [25]) and the rest of the wings. Each pMSP component was usually present in a higher amount in a single androconial unit considered as actively releasing that component. Lines represent the percentage of the pMSP components assigned to each unit per species; blank cells

correspond to androconial units absent in the species. Grey cells pinpoint androconial units containing 50% or more of the pMSP components found in the species. The last two columns show the total number of androconial units and the proportion active. The last line shows the proportion of all pMSP components present primarily in each unit. **b:** Scheme showing the location of androconial units on wings.

Note S8: Effect of the environment on the diversity of pMSP composition

We investigated the effect of habitat type (savannah, forest) on the diversity of pMSP composition using two approaches in addition to those described in the main text. First, we analysed the number of pMSP components per species as a function of habitat by using generalised linear models (GLM) with a Poisson distributed error. Since species diversity is higher in the forest habitat, we also included the number of sympatric species as a factor in the analysis. Model selection was based on corrected Akaike information criterion [86] (AICc) and the best model was:

number of pMSP components ~ number of sympatric species + habitat + number of sympatric species : habitat.

Since some of the effects were significant, we corrected for phylogenetic non-independence of species [87]. For this purpose we used the generalized estimating equations [88] (GEE) in the R package APE [38,89] (version 2.13.1 and 2.5-3 respectively). Results showed that forest species have a higher number of pMSP components than savannah species (effect size = 4.52; $P < 0.01$; ESM figure S13 and table S3). We also observed a significant interaction between habitat type and the number of sympatric species, which is difficult to explain: in forest habitat, the higher species diversity is associated with fewer pMSP components per species, while the opposite is true in savannah habitats. As these results were obtained using only three locations per habitat, this effect must be taken with caution and should be checked by sampling a larger number of locations per habitat.

Second, we examined the effect of habitat on pMSP composition by plotting the number of pMSP differences for pairs of species that either shared, or did not share, the same habitat. Species of different habitats did not differ more in pMSP composition than those sharing the

same habitat, indicating a limited effect of habitat on pMSP differentiation at this geographic scale (ESM figure S14). Most pMSP differentiation is present within the forest habitat, which was to be expected given that forest species have more pMSP components overall (ESM figure S14 and table S3). Because habitat affects the number of pMSP differences between species, habitat was included as a factor in the main analysis of RCD on pMSP composition (figure 3). The coding was ‘0’ for savannah pairs of species, ‘1’ for mixed pairs of species, and ‘2’ for forest pairs of species.

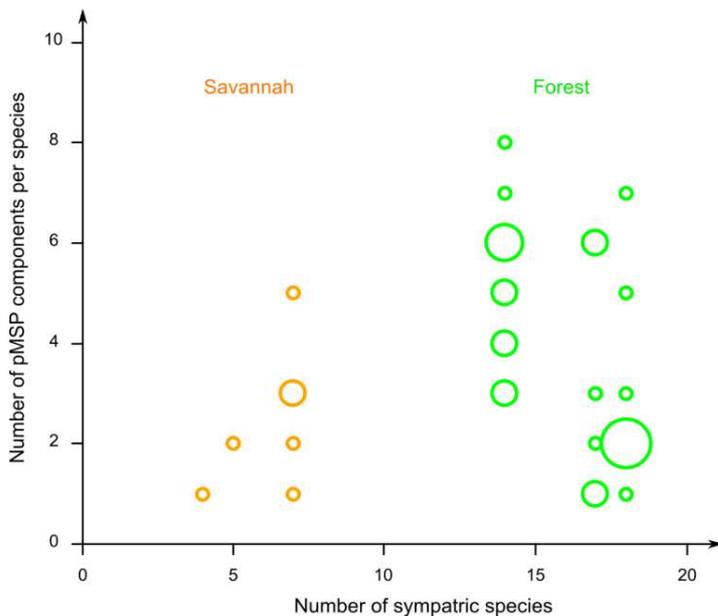


Figure S13: Number of pMSP components per species as a function of habitat and the number of sympatric species surrounding the focal species.

The size of the circles is proportional to the number of species (1 to 4). Savannah species are in orange and forest species in green.

Table S3: Best fit model parameters and their significance for the generalised estimating equations.

Factors	Estimates	s.e.	t	<i>P</i>
Intercept	-0.68	1.16	-0.59	0.57
Number of sympatric species	0.30	0.16	1.87	0.09
Habitat (Forest)	4.52	1.21	3.75	< 0.01
number of sympatric species : habitat:	-0.45	0.16	-2.77	0.02

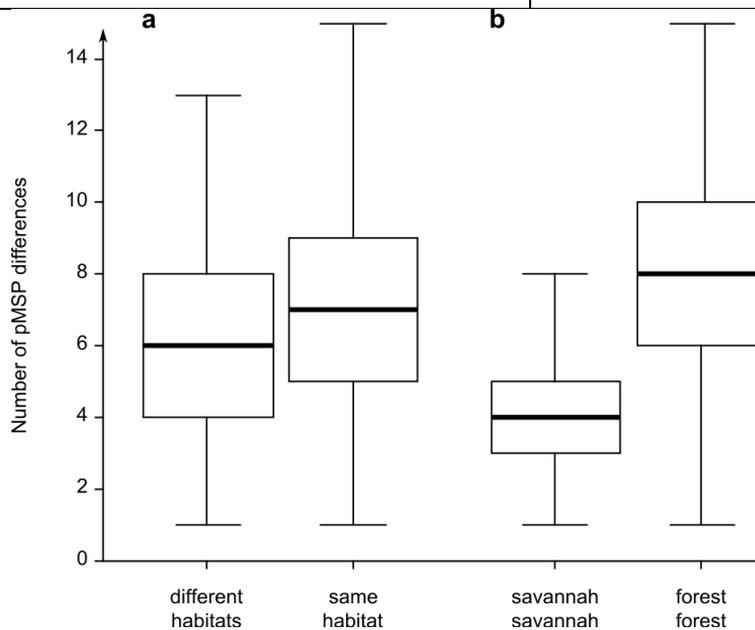


Figure S14: Number of pMSP differences based on the habitat occupied by the pairs of species.

a: Box-plots of the number of pMSP differences for pairs of species that either live in different habitats or share the same habitat. **b:** Box-plots of the number of pMSP differences for pairs of species living both in savannah or both in forest habitat (subsets of the second category of panel a).

Note S9: Sympatry and age of the pairs of species - the differential fusion hypothesis

We show that species sampled in allopatry are not more closely related than species sampled in sympatry (ESM figure S15; one tailed permutation test; 9999 permutations, $P = 0.93$). This type of test should be interpreted with caution since species ranges can change over time [90].

Moreover, our assessment of sympatry only depends on the location where species were sampled, not on sympatry at the scale of the entire range of the species. In any case, this pattern is not in accordance with the differential fusion hypothesis [49] which predicts that allopatric species are more closely related than sympatric species, since sympatric species are the subset of allopatric species that were differentiated enough to coexist after secondary contact [12,42].

Of note, we obtained pMSP data for most sympatric species in each of four sampled locations during extensive field work, which limits the risk of underestimating the relative age of sympatric, compared to allopatric, pairs of species. Moreover, in our sample of 32 *Bicyclus* species, we found only two cases where the original taxonomic description [24] had missed cryptic species living in sympatry, namely for *B. mesogena* in Uganda and *B. mandanes* in Cameroon. These taxonomic changes reinforce our results as it reduces further the average phylogenetic distance between pairs of sympatric species.

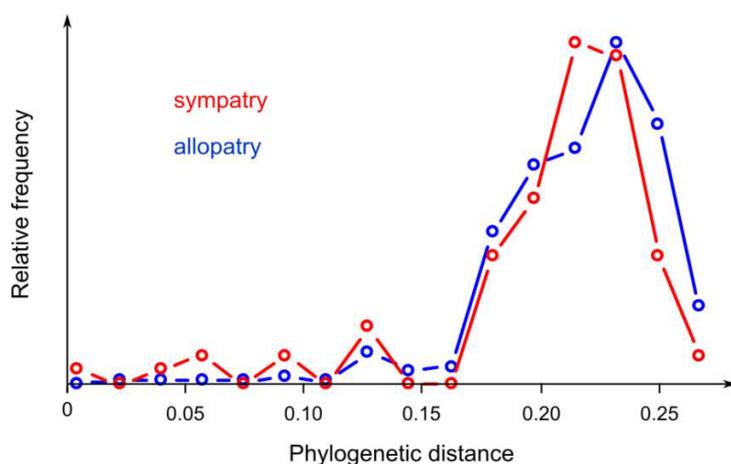


Figure S15: Distribution of the phylogenetic distance between species depending on their sympatric or allopatric status in our study.

The distances are based on the Bayesian tree (ESM note S2). The frequencies (ordinate axis) are scaled between the two groups (90 sympatric and 406 allopatric pairs of species).

Authors' contributions

P.M.B.B., O.B., P.M.B. and C.M.N. conceived and designed the study. The pheromone field sampling protocol was developed and tested by C.L., H.-L.W. and O.B. O.B. did the main data collection in the field (Liberia, Nigeria and Uganda) with the contribution of P.M.B.B. (Cameroon and Uganda) and C.M.N. (Cameroon). H.-L.W. performed chemical analyses, and H.-L.W. and C.L. identified compounds. P.M.B.B., O.B. and H.-L.W. did the post treatment of raw chemical data. O.B. conducted the scoring of androconia. P.M.B.B. did molecular analyses, carried out the phylogenetic and statistical data analyses. P.M.B.B. and C.M.N. wrote the manuscript with the contribution of O.B., H.-L.W., C.E.A., C.L. and P.M.B.

Supporting references

61. Brakefield, P. M., Beldade, P. & Zwaan, B. J. 2009 The African butterfly *Bicyclus anynana*: a model for evolutionary genetics and evolutionary developmental biology. *CSH Prot.* **2009**, pdb.emo122. (doi:10.1101/pdb.emo122)
62. Condamin, M. 1968 Mises au point de synonymie et descriptions de nouveaux *Bicyclus* (Lepidoptera satyrinae). *Bull. I.F.A.N.* **30**, 599–605.
63. Dunkelblum, E., Tan, S. H. & Silk, P. J. 1985 Double-bond location in monounsaturated fatty acids by dimethyl disulfide derivatization and mass spectrometry: Application to analysis of fatty acids in pheromone glands of four lepidoptera. *J. Chem. Ecol.* **11**, 265–277. (doi:10.1007/BF01411414)
64. Miller, L. 1969 Nomenclature of wing veins and cells. *J. Res. Lepid.* **8**, 37–48.
65. Robertson, K. A. & Monteiro, A. 2005 Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *Proc. R. Soc. Lond. B* **272**, 1541–1546. (doi:10.1098/rspb.2005.3142)
66. Costanzo, K. & Monteiro, A. 2007 The use of chemical and visual cues in female choice in the butterfly *Bicyclus anynana*. *Proc. R. Soc. Lond. B* **274**, 845–851.
67. Olofsson, M., Vallin, A., Jakobsson, S. & Wiklund, C. 2010 Marginal eyespots on butterfly wings deflect bird attacks under low light intensities with UV wavelengths. *PLoS ONE* **5**, e10798. (doi:10.1371/journal.pone.0010798)
68. Monteiro, A. & Pierce, N. E. 2001 Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII, and EF-1 alpha gene sequences. *Mol. Phyl. Evol.* **18**, 264–281.
69. Guindon, S. & Gascuel, O. 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
70. Posada, D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
71. Huelsenbeck, J. P. & Ronquist, F. 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
72. Ronquist, F. & Huelsenbeck, J. P. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
73. Altekar, G., Dwarkadas, S., Huelsenbeck, J. P. & Ronquist, F. 2004 Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**, 407–415. (doi:10.1093/bioinformatics/btg427)

74. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)
75. Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A. & Winka, K. 2003 Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* **52**, 477–487. (doi:10.1080/10635150390218213)
76. Zwickl, D. J. 2006 Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.
77. Bazinet, A. L. & Cummings, M. P. 2011 Computing the Tree of Life: leveraging the power of desktop and service grids. In *Proceedings of the 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum*, pp. 1896–1902. Washington, DC, USA: IEEE Computer Society. (doi:10.1109/IPDPS.2011.344)
78. Maddison, W. P. & Maddison, D. R. 2011 *Mesquite: a modular system for evolutionary analysis. Version 2.75*.
79. Mooers, A., Vamosi, S. & Schluter, D. 1999 Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). *Am. Nat.* **154**, 249–259.
80. Kluge, A. G. & Farris, J. S. 1969 Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**, 1–32.
81. Farris, J. S. 1989 The retention index and homoplasy excess. *Syst. Zool.* **38**, 406–407. (doi:10.2307/2992406)
82. Schliep, K. P. 2011 Phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593. (doi:10.1093/bioinformatics/btq706)
83. Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P. & Webb, C. O. 2010 Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464. (doi:10.1093/bioinformatics/btq166)
84. Sanderson, M. J. & Donoghue, M. J. 1989 Patterns of variation in levels of homoplasy. *Evolution* **43**, 1781–1795. (doi:10.2307/2409392)
85. Phelan, P. L. & Baker, T. C. 1987 Evolution of male pheromones in moths: reproductive isolation through sexual selection. *Science* **235**, 205–207.
86. Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. Second edition. New York, NY, USA: Springer.
87. Abouheif, E. 1999 A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* **1**, 895–909.

88. Paradis, E. & Claude, J. 2002 Analysis of comparative data using generalized estimating equations. *J. Theor. Biol.* **218**, 175–185.
89. Paradis, E., Claude, J. & Strimmer, K. 2004 APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290. (doi:10.1093/bioinformatics/btg412)
90. Losos, J. B. & Glor, R. E. 2003 Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.* **18**, 220–227. (doi:10.1016/S0169-5347(03)00037-5)