Drivers of diversification in *Heliconius*, with special focus on the *sara/sapho* clade

NICOL MAGALY RUEDA MUÑOZ
DOCTORATE PROGRAM IN BIOMEDICAL AND BIOLOGICAL SCIENCES
UNIVERSIDAD DEL ROSARIO

This dissertation is submitted for the Degree of Doctor of Philosophy
Junio 2023

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text or the acknowledgements below. None of the work has been submitted for a degree, diploma, or dissertation at any other university.

Nicol Magaly Rueda Muñoz

SUMMARY

Understanding the mechanisms and processes driving biological diversification and adaptation is still a major question in evolutionary biology that requires interdisciplinary research that addresses the role of biotic (i.e., genetic background, ecological interactions) and abiotic factors (i.e., climate). In this dissertation I studied biogeographic, chromosomic, and chemical aspects that contribute to the diversification of the Neotropical butterflies of the genus *Heliconius*, especially species in the *sara/sapho* clade. Although *Heliconius* is one of the best studied groups in the context of evolutionary biology and ecology, the clade *sara/sapho* has been largely unstudied despite having unique features. For example, some of its species show high diversification rates and a higher number of chromosomes compared to other *Heliconius*, and also, species in the clade seem unable to synthesise cyanogens leading to reliance on toxins sequestered from larval host plants.

In Chapter I, I used 54,392 georeferenced records for 46 species and 1,012 georeferenced records for 38 interspecific hybrids of *Heliconius* to investigate the role of the environment in shaping their distribution and richness, as well as their geographic patterns of phylogenetic diversity and phylogenetic endemism. I also evaluated whether niche similarity promotes hybridization. I found that *Heliconius* displays five general distribution patterns mostly explained by precipitation and isothermality, and to a lesser extent, by altitude. Interestingly, altitude plays a major role as a predictor of species richness and phylogenetic diversity, while precipitation explains patterns of phylogenetic endemism. I did not find evidence supporting the role of the environment in facilitating hybridization because hybridizing species do not necessarily share the same climatic niche despite some of them having largely overlapping geographic distributions. Overall, I confirmed that, as in other organisms, high annual temperature, a constant supply of water, and spatial-topographic complexity are the main predictors of diversity in *Heliconius*.

In Chapter II, I generated whole genome resequencing data for 114 individuals from seven species in the *sara/sapho* clade to investigate: (i) genome-wide phylogenetic relationships, (ii) the degree of genomic differentiation between species and subspecies, and (iii) the impact of chromosomal rearrangements in the evolution of the clade. The inclusion of multiple species and subspecies of this clade allowed me to redefine some of the relations previously reported, and to identify the effect of geography in shaping their diversity. Interestingly, I also found evidence for sex-autosome fusions involving autosomes 4, 9, and 14. All of these fusions seem to be associated with speciation events in this clade, with the sex-autosome 4 fusion being the oldest one. Although I do not yet understand the role or evolutionary consequences of these fusions, my study shows that chromosomal rearrangements can evolve rapidly within a clade and generate chromosomal diversity.

In Chapter III, I investigated how cyanogens (chemical defences of adult *Heliconius*) vary both in composition and concentration across nine mimicry rings and six Neotropical ecoregions. I found that variation in the cyanogenic profile of *Heliconius* is not explained by the mimicry ring that a species belongs to or its locality. Instead, cyanogenic variation is the result of phylogenetic closeness and, likely, ecological factors such as host plant specialization, diversity and abundance of local hostplants locally available, availability of precursors for biosynthesis of cyanogenic compounds in pollen-source plants, as well as the local predator community. My results agree with recent modelling and meta-analyses that showed that increased toxicity of preys does not translate into increased predator learning or generation of mimetic diversity.

ACKNOWLEDGEMENTS

Este trabajo no habría sido posible sin la supervisión del Dr. Camilo Salazar y la Dr. Carolina Pardo. A ellos agradezco sus innumerables enseñanzas, su actitud siempre dispuesta para ayudarme a sortear los momentos difíciles de este proceso, y por propiciar increíbles experiencias académicas y de campo.

Quiero agradecer al Ministerio de Ciencias de Colombia por la beca asignada para mi matrícula y sostenimiento (Convocatoria 727 de 2016), así como a la Universidad del Rosario por facilitar los recursos para mi pasantía en la Universidad de Cambridge (Apoyo para Estudiantes Doctorales), y al departamento de Biología por su contribución económica durante la pandemia.

Quiero agradecer al Dr. Chris Jiggins y la Dr. Joana Meier por la supervisión de mi trabajo durante la pasantía en la Universidad de Cambridge. En particular, quiero agradecer a Joana Meier por acompañarme en mis primeros pasos en el campo de la genómica. Gracias por tanta paciencia y enseñanzas.

Agradezco a todas las personas que participaron y colaboraron en la culminación de este trabajo: Fabian Salgado, Carlos Gantiva, Frasella de Martino, Alexandra Arias, Dr. Carol Garzón, Dr. Carlos Arias, Dr. Gabriela Montejo-Kovacevich, Dr. Érika C. Pinheiro de Castro y Dr. Mónica Arias.

Fue un gusto trabajar con mis compañeros del Grupo de Genética Evolutiva de la Universidad del Rosario. En particular, quiero agradecer a Fabian Salgado, Melissa Sánchez y Andrea Thomaz por el apoyo académico y emocional.

A mis compañeros de doctorado y amigos Carolina Álvarez, Carlos Vargas, Carlos Gantiva y Oswaldo Gil quiero agradecer por tantas charlas llenas de historias divertidas y por el apoyo mutuo.

Agradezco a todos los amigos que me apoyaron de una u otra forma durante estos años. Las charlas, consejos y risas de Nancy Jerez, Carolina Hernández y Angelica Tipan hicieron más fácil este proceso.

No hay palabras suficientes para agradecer a mi familia. A mis padres por tan infinito amor, tanta dedicación y por haberme enseñado la importancia del trabajo y la responsabilidad. Sus consejos me han permitido llegar hasta aquí. A Oscar, Catalina y Jaime Reyes por su apoyo incondicional. A Geraldine por sus consejos, alegría y amor. A mis sobrinos Luciana y Martin por tantos besos, abrazos y momentos de felicidad que me motivaron a continuar. Y a mi esposo Jaime Reyes mil y mil gracias. Sin ti nada de esto sería posible. Gracias por tu apoyo económico y emocional, gracias por caminar a mi lado, por creer en mí y darme ánimos cuando más lo necesitaba.

COLLABORATIONS AND PUBLICATIONS

This thesis is the result of my work in collaboration with other colleagues.

I obtained data from Chapter I by visiting the following collections in Colombia: (i) Institute of Natural Sciences - Universidad Nacional (Bogotá), (ii) Institute for the Research of Biological Resources Alexander von Humboldt, (iii) Universidad de los Andes, (iv) Universidad Nacional (Medellín), (v) Private collection of Jean Francois Le Crom, and (vi) Universidad del Rosario. I also obtained data from Neil Rosser, who was very open about sharing his occurrence records. Additional fieldtrips to regions poorly sampled were conducted by me, Dr. Camilo Salazar, Dr. Carolina Pardo and Carlos Gantiva. I curated the databases with the help of Carlos Gantiva. I performed all data analyses, with some help of Fabian Salgado. I wrote the manuscript under the supervision of Dr. Camilo Salazar and Dr. Carolina Pardo. The result of this Chapter was published in *Frontiers in ecology and evolution* (Q1) where I am first author:

> **Nicol Rueda-M**, Fabian C Salgado-Roa, Carlos H Gantiva-Q, Carolina Pardo-Diaz, Camilo Salazar (2021). *Environmental drivers of diversification and hybridisation in Neotropical butterflies*. Frontiers in Ecology and Evolution. doi: 10.3389/fevo.2021.750703

Part of the data of this chapter were also published as a book chapter where I am the first author:

> **Nicol Rueda-M**, Carlos H. Gantiva-Q, Miguel G. Andrade-C (2021). *Mariposas de la reserva Natural Bojonawi y zonas adyacentes (Escudo Guayanés), Orinoquia, Vichada, Colombia*. In: VII. Biodiversidad de la reserva natural Bojonawi, Vichada, Colombia: rio orinoco y planicie de inundación. Instituto Humboldt. ISBN digital: 978-958-5183-02-5.

In Chapter II, the reference genome of *Heliconius sara* was generated in the Sanger Institute by Drs. Chris Jiggins, Shane McCarthy and Richard Durbin. I, Dr. Camilo Salazar and Dr. Carolina Pardo collected the samples of the *sara/sapho* clade in Colombia. Samples from other countries were collected by Drs. Chris Jiggins, Owen McMillan, Krzysztof M. Kozak, Jonathan Ready, and Carlos Arias. I performed all data analyses under the supervision of Dr. Joana Meier and Dr. Camilo Salazar, and I got help from Dr. Gabriela Montejo-Kovacevich with mapping the reads of the first individuals to the reference genome. I wrote the manuscript with the help of Drs. Camilo Salazar, Joana Meier, and Dr. Carolina Pardo. This chapter is published as a preprint in bioRxiv, and it went through the first round of revisions *PLoS Genetics*:

> **Nicol Rueda-M**, Chris D. Jiggins, Carolina Pardo-Diaz, Gabriela Montejo-Kovacevich, W. Owen McMillan, Shane McCarthy, Jonathan Ready, Krzysztof M. Kozak, Carlos F. Arias, Richard Durbin, Joana I. Meier, Camilo Salazar (2023). Three sequential sex chromosome – autosome fusions in *Heliconius* butterflies. bioRxiv 2023.03.06.531374; doi: https://doi.org/10.1101/2023.03.06.531374

In Chapter III, I, Dr. Camilo Salazar, and Dr. Carolina Pardo collected the samples of the *sara/sapho* clade in Colombia. Sample pre-processing was conducted by me and Dr. Érika C. Pinheiro de Castro, and she ran the Analytical LC-MS. I performed all data analyses supervised by Drs. Camilo Salazar, Monica Arias and Érika C. Pinheiro de Castro. I wrote the manuscript with the help of Drs. Camilo Salazar, and Dr. Carolina Pardo. This chapter was written in view of eventual publication and is under preparation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.

Environmental drivers of diversification and hybridization in neotropical butterflies

Introduction

Understanding how the environment shapes species distribution and affects patterns of biological diversity is still a challenging task, especially in species rich regions such as the Neotropics (1–3). To date, information on this topic is mostly based on vertebrates and plants, and suggest that the combination of high annual temperature with a constant supply of water and spatial-topographic complexity are the main predictors of species distribution, richness and endemism (4–7). Within the Neotropics, the Amazon and the foothills of the North-eastern Andes are examples of regions that combine these conditions, and consequently, they exhibit high levels of species richness and phylogenetic diversity in monkeys, snakes, birds, amphibians, palms and vascular plants (8–11). Similarly, regions such as the Biogeographic Choco, Costa Rica and the Amazon show high levels of phylogenetic endemism (12–14). However, these patterns have not been deeply evaluated in Neotropical invertebrates, and particularly butterflies (15,16).

The environment, and especially climatic niche, has also been suggested to have an effect on gene flow. For example, phylogenetic discordance in multiple loci in beetles of the genus *Mesocarabus* seems to be the result of hybridization between species sharing the same climatic niche (17), while in armadillos of the genus *Dasypus*, asymeytric gene flow appears to be facilitated by niche conservatisms at both sides of a geographic barrier (18). Additionally, climatic-based selection likely plays a role in maintaining mosaic hybrid zones in *Quercus* oaks, where climatic heterogeneity favors the co-occurrence of parental species and their hybrids (19,20). *Heliconius* butterflies are a diverse insect group found across southern United States, Central and South America, where they occupy divergent habitats (21). Due

to the recent radiation of this butterfly genus, species pairs have different levels of reproductive isolation, which are used as proxies for different stages of speciation (22,23). In total, ~25% of *Heliconius* species are known to hybridize in nature (24,25), but the role of abiotic variables in facilitating or hampering such hybridization has been poorly studied (26,27).

In this study, we combined an extensive database of occurrences of species and hybrids in *Heliconius* as well as environmental data to investigate: 1) how the environment shapes the distribution of *Heliconius* at a regional scale, 2) how the environment molds species richness, phylogenetic diversity, and phylogenetic endemism in these butterflies, and 3) whether niche similarity promotes hybridization.

Materials and Methods

Species data and environmental variables

We included occurrence data of 46 species of *Heliconius* and generated a database of the localities where these butterflies have been collected across their entire distribution range. The data were obtained from: (1) entomological collections, and (2) the Heliconiinae checklist of (28). For those regions in Colombia that we identified as under-sampled, we conducted field trips to improve our geographic coverage. The nomenclature of all records was updated to the most recent taxonomic checklist when needed (29). We also included occurrence data for all interspecific hybrids documented in *Heliconius.* All individuals were photographed and identified based on their color pattern. We used the point-radius method to georeference specimens with missing coordinates following (30). Although *Heliconius* is widely represented in databases such as GBIF, we did not include such records to ensure the use of data that have been curated by specialists both in terms of georeference and taxonomy, or that have images of each specimen that would allow us to confirm the taxonomy.

We used the 19 climatological variables from CHELSA at spatial resolution of 1 Km (31) to characterize climatic variation across the occurrence range of *Heliconius*, and altitude was obtained from (32). Collinearity between variables was avoided by estimating the Pearson correlation coefficient among all 20 variables, and the absolute value of this correlation was used to create a dissimilarity matrix (1-correlation values). We used this matrix was used to perform a hierarchical clustering analysis with the *hclust* function in R (33). We then chose one variable per cluster that had a pairwise distance less than 0.5. Using the selected variables, we calculated the variance inflation factor (VIF) (34) with the *HH* package in R (35) and chose those variables with VIF less than 5 (36).

Species distribution modelling and environmental variables importance

First, we used R pipelines (37) to reduce sampling bias and spatial autocorrelation among occurrences in our species distribution models using the variables that passed the filters mentioned before. The minimum non-significant autocorrelated distances were used to prune species databases. *H. nattereri* and *H.tristero* were not modeled because they had less than 32 occurrence records. Then, we generated a second database that included pseudo-absences data following (38–41). Because *Heliconius* is a very well sampled genus we had enough information to select pseudo-absences points for each species in places where: (i) *Heliconius* other than the focal species have been collected, (ii) environmental conditions may not be optimal for its occurrence, and (iii) absence is not caused by dispersal limitation. Using these criteria, we defined a minimum convex polygon with a 50 Km buffer area for each species and selected 10,000 pseudo-absences only in this buffer.

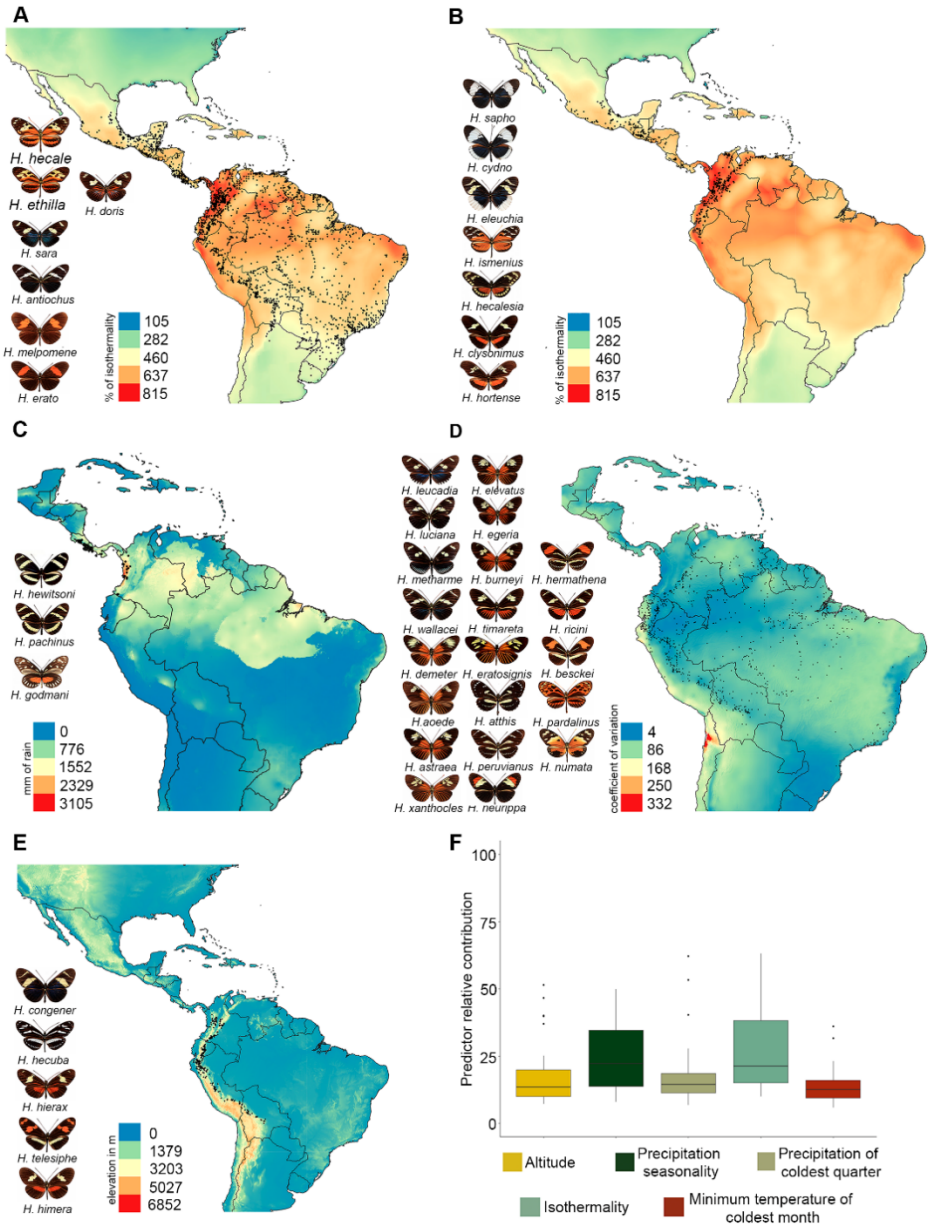

Figure 1.1 Distribution patterns of *Heliconius* species based on environmental variables. (A) Species with wide distribution (explained by isothermality); (B) species with trans-Andean distribution (explained by isothermality); (C) species distributed in the biogeographic Choco + Costa Rica (explained by precipitation of coldest quarter); (D) species distributed in the cis-Andes + Pacific of Ecuador (explained by precipitation seasonality); (E) species distributed in highlands of the Andes (explained by altitude); (F) relative importance of environmental variables receiver operating characteristic (ROC) that are predictors of diversity in *Heliconius*. Colour scale in panels (A–E) indicates the variable gradient. Distribution maps for each of the species can be found at: https://doi.org/10.5281/zenodo.5149294.

Then, we estimated the ensemble species distribution models (ESDMs) of *Heliconius* with the R package SSDM (42), equally weighting presences and pseudo-absences (prevalence weights = 0.5). Individual SDM models were implemented using four algorithms that optimize the use of pseudo-absences in a similar way (38): (1) Generalized Linear Models (GLM) (43), (2) Generalized Boosting Models (GBM) (44), (3) Maximum Entropy Models (MAXENT) (45), and (4) Generalized additive model (GAM) (46). Each algorithm was run 10 times. In each run, models were calibrated using 75% of the occurrence data and their accuracy was evaluated with the remaining 25%; the "holdout" method was used to ensure independence between training and evaluation sets. The data set randomly changes between runs. An ensemble model (ESDM) was obtained for each species by averaging the best SDM outputs (highest Area Under the Curve – AUC - score), and the ensemble models were evaluated with the AUC score and the Cohen's Kappa coefficient (k). Following (47), we did not model species with n <32 or that occupy >70% of the background region (i.e., entire distribution range for the genus).

We used the relative importance values of the variables provided by SSDM to evaluate the influence of each of them within all models. The importance is estimated with a randomization process, where SSDM calculates the correlation between a prediction using all variables and a prediction where the independent variable being tested is randomly removed; this is repeated for each variable. The calculation of the relative importance is made by subtracting this correlation from one, therefore higher values are the best variables for the model (48).

Diversity metrics: species richness, diversity, endemism phylogenetic maps and environmental variables importance

Species richness, phylogenetic diversity and phylogenetic endemism were calculated by superimposing the distribution maps of all species using the R package phyloregion (49). In order to avoid overestimation of the diversity metrics, we created alpha hulls with the R package rangeBuilder (50) and following (51). Briefly, we used occurrence data available for all species (54,392 georeferenced records) that had

more than ten locality points, a dynamic selection of alpha for each species, and an alpha that varied in steps of 1 (52). We next generated a community matrix using the alpha hulls of all species with the function polys2comm in the R package phyloregion (53).

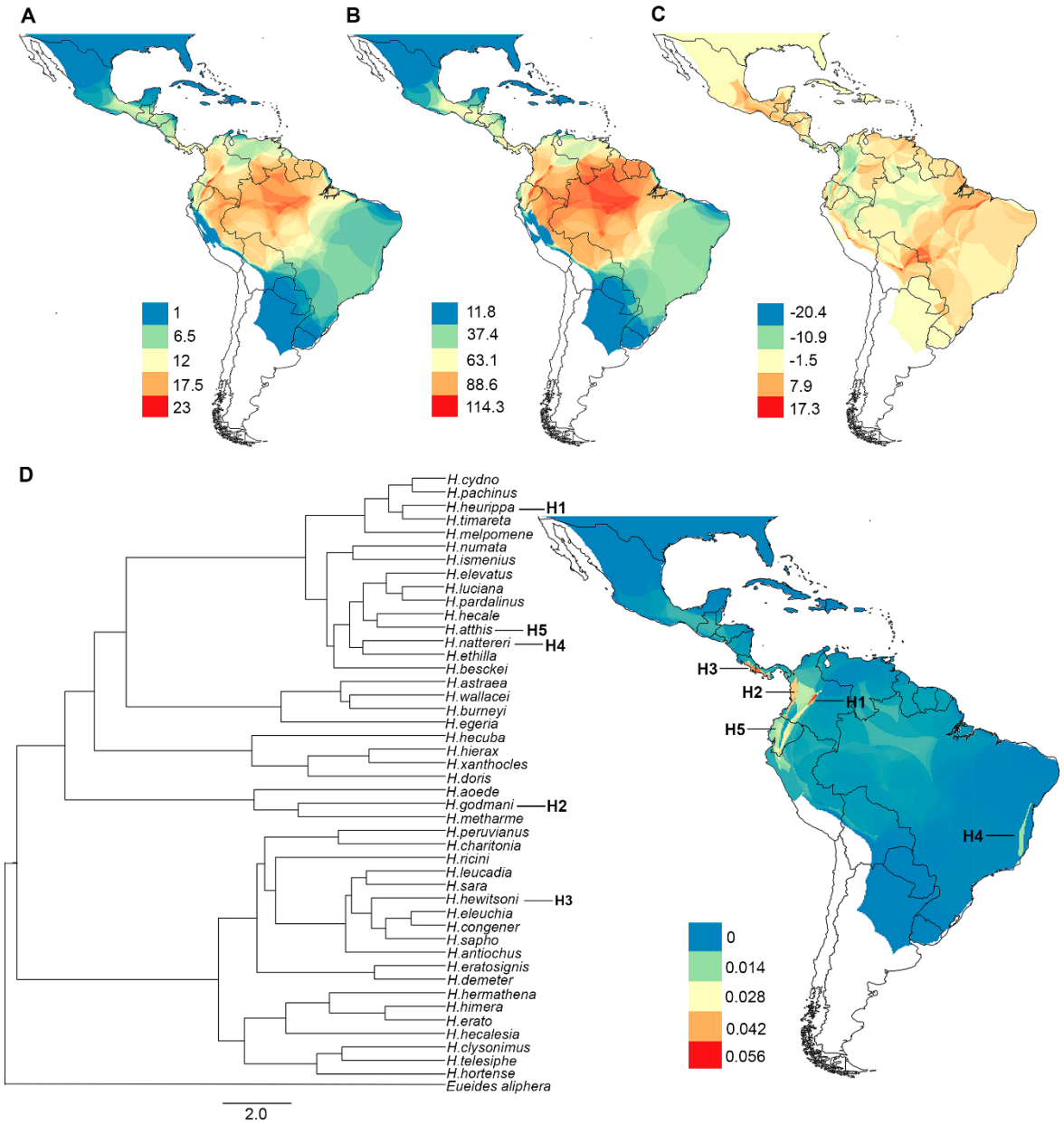

Figure 1.2 Maps of diversity metrics. (A) Species richness, (B) phylogenetic diversity, (C) residuals of phylogenetic diversity regressed on species richness, and (D) phylogenetic endemism. Warm colours indicate higher values, while cold colours are indicative of lower values. The phylogeny shown in panel (D) was modified from Kozak et al. (2015), and

branches that contribute the most to the phylogenetic endemism are labelled as H1–H5, both in the phylogeny and the map. All maps were plotted in grid cells of 20 Km x 20 Km.

We used the community matrix to calculate species richness by summing all species present in each cell, and also, with this matrix and the best Maximum Likelihood tree estimated with 20 nuclear and 2 mitochondrial loci for *Heliconius* (54), we estimated phylogenetic diversity and phylogenetic endemism (55,56), with the functions *PD* and *phylo_endemism* of the R package *phyloregion* (53). To investigate whether these metrics are scale dependent, we performed the above analyses at three consecutive grain sizes (5Km, 10Km and 20Km). We performed a linear regression model using phylogenetic diversity as response variable and species richness as predictor variable to investigate their relationship and plotted the residuals to highlight areas where these metrics are different.

We also used four machine learning algorithms to generate correlative models and then we created an ensemble prediction of each diversity metric to identify the environmental variables that best explain them (57). The algorithms used were: Random Forests (58), Neural Network (59), Support Vector Machines (60), and Generalized Linear Model (43). The models were built with the R package *caret* 6.0-86 (61), and we used the *varImp* function to compute the weighted average of the contribution of each variable.

Evaluating the environmental effect in the hybridization on Heliconius butterflies

We estimated the Schoener's niche equivalency test (D) and Warren's niche background test (I) between pairs of hybridizing species to determine if they share environmental niches. We used the R package *humboldt* (62) and we followed the concept of environmental niche *sensu* (45,63), where the niche consists of the subset of conditions currently occupied and where environmental conditions at the occurrence localities constitute samples from the realized niche. The niche overlap metric *Schoener's D* ranges between 0 and 1, meaning no overlap and complete overlap, respectively (64). The environmental overlap was visualized with a PCA. We tested the significance of this metric by comparing the realized niche overlap

against a null distribution of 1,000 overlaps randomly generated from the reshuffled occurrence dataset and tested whether niche background and niche equivalency were different from those expected by chance at $\alpha = 0.05$ (62). This was done using the entire distribution of the entities under comparison (niche overlap test = NOT) and using only the area where they overlap (niche divergence test = NDT) (62).

Results

Species data, species distribution modelling and environmental variables importance

We collected a total of 68,877 records for 46 species (n=67,865), 37 cases interspecific hybrids (n=164), and 34 cases of intraspecific hybrids (n=848) in *Heliconius* (Table S1.1 and Table S1.2). From the species records we discarded 13,476 records as they could not be reliably georeferenced, thus leaving us with 54,392 records. For species modeling, these were further subject to pruning, which left a total of 13,661 records (Table S1.3). There was considerable variation in the sampling effort across the phylogeny. For example, species of the erato and silvaniform clades are well represented, whereas species from the aoede clade had the lowest number of records (Figure S1.1). The variables retained and used to model species distributions and diversity metrics were: (i) min temperature of coldest month, (ii) altitude, (iii) precipitation of coldest quarter, (iv) isothermality, and (v) precipitation seasonality (Figure S1.2). The maximum absolute pairwise correlation between min temperature of coldest month and precipitation of coldest quarter was 0.436. The four algorithms we implemented were accurate in predicting the distribution of species, but their combination (ensemble) was the most accurate (Figure S1.3). In total, we generated 44 species distribution models for *Heliconius* species. These are deposited in ZENODO (https://doi.org/10.5281/zenodo.5149294).

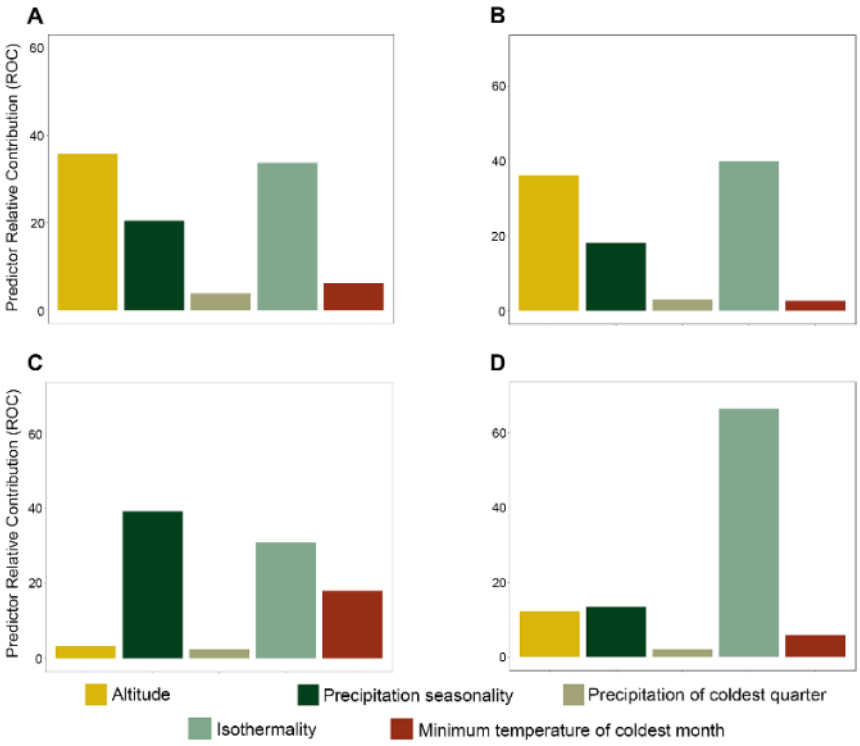

Figure 1.3 Relative importance of predictors (environmental variables) of diversity. (A) Species richness, (B) phylogenetic diversity, (C) phylogenetic endemism, and (D) residuals of the phylogenetic diversity/species richness regression.

We found that environmental variables are better predictors of the distribution of *Heliconius* compared to topography. For instance, current temperature (isothermality) explains the distribution of 14 species (Figure 1.1A, B) and precipitation explains the distribution of 24 species (Figure 1.1C, D). In contrast, altitude explains the distribution of only five species (Figure 1.1E). No single variable was correlated with the entire distribution of the genus (Figure 1.1F), but we observed some general patterns. For example, isothermality explained the distribution of widely distributed species and trans-Andean species (i.e., west of the Andes; Figure 1.1A, B). Also, precipitation of the coldest quarter explains the distribution of species that occur in the biogeographic Choco + Costa Rica while precipitation seasonality explains the distribution of cis-Andean species (i.e., east of the Andes) + the Pacific of Ecuador (Figure 1.1C, D). Altitude explains the distribution of species restricted to the eastern foothills of the Andes and highland

Andean species (Figure 1.1E). Interestingly, we did not find a single variable that was better correlated with the distribution of *H. charithonia* (Table S1.4).

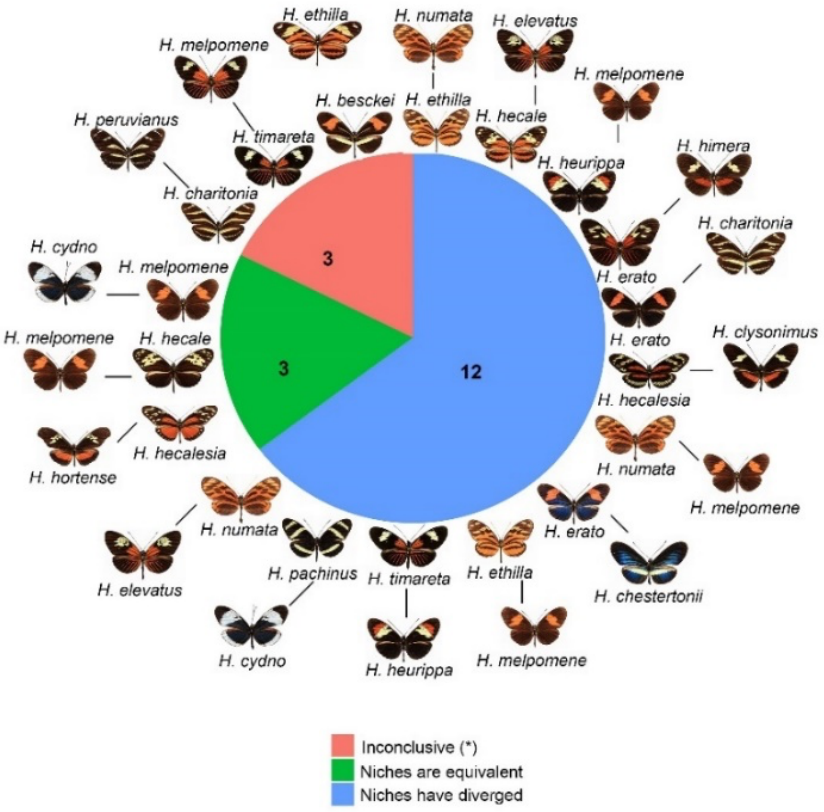

Figure 1.4 Co-occurring and hybridizing species of *Heliconius*. Green: species pairs with equivalent environmental niches, blue: species pairs with divergent environmental niches, and salmon: species pairs with inconclusive results. Numbers indicate the pairs of species falling into each category.

Diversity metrics: species richness, diversity, endemism phylogenetic maps and environmental variables importance

We found that higher values of *Heliconius* species richness are concentrated in the foothills of the eastern Andes from Colombia to Ecuador, and into the Amazon basin mainly along the course of the Amazon River (Figure 1.2A). These results were consistent but more striking in the phylogenetic diversity maps (Figure 1.2B). Also, species richness has a strong and significant effect on phylogenetic diversity (adjusted R^2 0.9887, p= <2e-16; Figure S1.4). Interestingly, the residuals map

showed values of phylogenetic diversity below those expected from species richness in the same regions, indicating that phylogenetic diversity, although high, is underestimated (blue grids; Figure 1.2C). In contrast, this metric was overestimated mainly in the Central Andes, the southern Amazon in Brazil, and the northern Chaco in Bolivia (red grids; Figure 1.2C). The highest values of phylogenetic endemism were concentrated in: (i) the Pacific coast of Costa Rica and Panama, (ii) the central foothills of the Eastern Cordillera in Colombia, and (iii) the biogeographic Choco of Colombia (Figure 1.2D). The pattern of these metrics was not scale dependent, and the results were highly congruent at 5Km, 10Km (Figures S1.5 and S1.6, respectively), and 20Km (Figure 1.2C).

The ability of the machine learning models to predict species richness, phylogenetic diversity, and phylogenetic endemism varied between algorithms (Figure S1.7). The best algorithms for all diversity metrics were the ensemble model followed by random forest, while the generalized linear model algorithm had the lowest predictive accuracy in all metrics (Figure S1.7). The best models predicted that altitude and isothermality were the most important variables for species richness and phylogenetic diversity (Figures 1.3A, B). In contrast, the most important variable for phylogenetic endemism was precipitation seasonality, followed by isothermality (Figure 1.3C). Finally, the residuals from the spatial regression between phylogenetic diversity (response variable) and species richness (predictor variable) were explained by isothermality (Figure 1.3D).

Evaluating the environmental effect on hybridization in Heliconius

We found 18 pairs of hybridizing species in *Heliconius.* The results of the NOT and NDT tests based on *Schoener's D* revealed that the niches of three of these pairs (*H. melpomene/H. cydno*, *H. melpomene/H. hecale*, and *H. hecalesia/H. hortense*) are equivalent (Figure 1.4; Table 1.1) and overlap climatically (D>0.40). In contrast, 12 of these pairs did not show evidence of niche equivalency. These included both pairs that have extensive geographic overlap (such as *H. ethilla* and *H. numata*) (Figure S1.8) and pairs with a narrow overlap (such as *H. erato* and *H. himera*)

(Figure 1.5). The remaining three pairs (*H. beskei/H. ethila*, *H. timareta/H. melpomene*, and *H. charitonia/H. peruvianus*) showed inconclusive results (Figure 1.4; Table 1.1). The results of these analyses were deposited in ZENODO (https://doi.org/10.5281/zenodo.5149294).

Table 1.1 Niche overlap test (NOT) and niche divergence test (NDT) between hybridizing species.

Species 1	Species 2	Niche overlap test (NOT)				Niche Divergence Test (NDT)				Interpretation
		Equivalency test		Background test		Equivalency test		Background test		
		D	p value for D	p value for D (2-1)	p value for D (1-2)	D	p value for D	p value for D (2-1)	p value for D (1-2)	
H. melpomene	*H. cydno*	0.4404	0.7924	0.0036	0.0026	0.4904	0.8643	0.0039	0.0027	**Niches are equivalent**
H. pachinus	*H. cydno*	0.1136	0.2255	0.1264	0.0114	0.1575	0.0100	0.0357	0.0132	Niches have diverged
H. ethilla	*H. numata*	0.3740	0.0020	0.0034	0.0154	0.3733	0.0020	0.0034	0.0182	Niches have diverged
H. melpomene	*H. ethilla*	0.2989	0.0020	0.0154	0.0026	0.3087	0.0020	0.0156	0.0027	Niches have diverged
H. besckei	*H. ethilla*	0.4262	0.9581	0.0149	0.2500	0.6228	0.8104	0.2000	0.1429	Inconclusive
H. melpomene	*H. heurippa*	0.0396	0.0120	0.0536	0.6728	0.0551	0.0020	0.0508	0.8587	Niches have diverged
H. elevatus	*H. numata*	0.3833	0.0579	0.0035	0.0556	0.3939	0.0379	0.0035	0.0526	Niches have diverged
H. timareta	*H. melpomene*	0.0903	0.7425	0.5182	0.0669	0.1106	0.3433	0.3485	0.0669	Inconclusive
H. melpomene	*H. hecale*	0.5138	0.1936	0.0033	0.0025	0.5243	0.2016	0.0038	0.0026	**Niches are equivalent**
H. hecale	*H. elevatus*	0.2655	0.0020	0.0385	0.0110	0.2731	0.0020	0.0333	0.0042	Niches have diverged
H. erato	*H. chestertonii*	0.0405	0.0019	0.6593	0.7804	0.0571	0.0001	0.6087	0.6926	Niches have diverged
H. erato	*H. charitonia*	0.0170	0.0020	0.0050	0.0021	0.2426	0.0100	0.0050	0.0022	Niches have diverged
H. charitonia	*H. peruvianus*	0.0172	0.7226	0.1429	0.8894	0.2404	0.9541	0.1250	0.0135	Inconclusive
H. hecalesia	*H. hortense*	0.4213	1.0000	0.0161	0.0130	0.4799	0.9940	0.0182	0.0323	**Niches are equivalent**
H. hecalesia	*H. clysonimus*	0.2464	0.0060	0.0032	0.0118	0.2359	0.0020	0.0032	0.0132	Niches have diverged
H. melpomene	*H. numata*	0.4600	0.0200	0.0036	0.0026	0.4843	0.0180	0.0034	0.0026	Niches have diverged
H. timareta	*H. heurippa*	0.2432	0.0998	0.0448	0.1111	0.1821	0.0020	0.0185	0.3103	Niches have diverged
H. erato	*H. himera*	0.0340	0.002	0.02632	0.0322	0.1100	0.0019	0.0333	0.0129	Niches have diverged

Discussion

We found that *Heliconius* butterflies display five general distribution patterns, namely: (i) wide distribution, (ii) trans-Andes, (iii) biogeographic Choco + Costa Rica, (iv) cis-Andes + Pacific of Ecuador, (v) highland Andes. We also found that three variables (isothermality, precipitation and altitude) explain these patterns. Isothermality is a variable that quantifies how daily temperatures oscillate relative to the annual oscillations (65), and its importance as one of the most explanatory variables of species distribution is not without precedent. For example, this variable explains the distribution of frugivorous bats (66), mealybugs (67), Opiliones (68), and American monkeys (8). Although all *Heliconius* species are strongly affected by isothermality, its effect seems to be stronger for widely distributed species and those with trans-Andean distribution. Interestingly, these species occur in regions with high and medium isothermality (>460%), that is, in regions that experience temperature changes throughout the day but keep a constant temperature throughout the year (65). This suggests that these butterflies are particularly sensitive to long term changes in temperature, thus limiting their range to tropical areas.

The distribution of species occurring in the biogeographic Choco of Colombia, cis-Andes and the Pacific of Ecuador is also strongly limited by precipitation seasonality. Consistently, these regions have either rainforest, monsoon, or savanna climate, and they are the Neotropical regions with the highest precipitation (precipitation in the driest month (Pdry) > 60 mm) (69). Previous studies have suggested that cloudiness and precipitation decrease flying bout duration in butterflies and, consequently, limit their dispersal (70). Therefore, exceptionally high levels of precipitation in such regions may act as population traps, preventing butterflies from flying over longer distances and keeping them in a single region (27). This finding agrees with previous studies in South America, where precipitation shapes the distribution of multiple vertebrates and invertebrates (71–74).

In addition, altitude was the best predictor for the distribution of *Heliconius* species that can reach elevations up to 2600 masl, which is considerably higher than the elevational range occupied by other members of the genus (<2200) (28).

Therefore, it is likely that these highland species have morphological or physiological modifications that allow them to expand their elevational range and occupy new niches. In fact, highland *Heliconius* are known to have rounder wings compared to lowland species, and this has been suggested to aid them flying dense cloud forest or compensate for the lower air pressure found at higher altitudes (75). Also, comparisons among different populations of *Heliconius* have revealed that highland populations are less tolerant to heat (76), which may limit their distribution range.

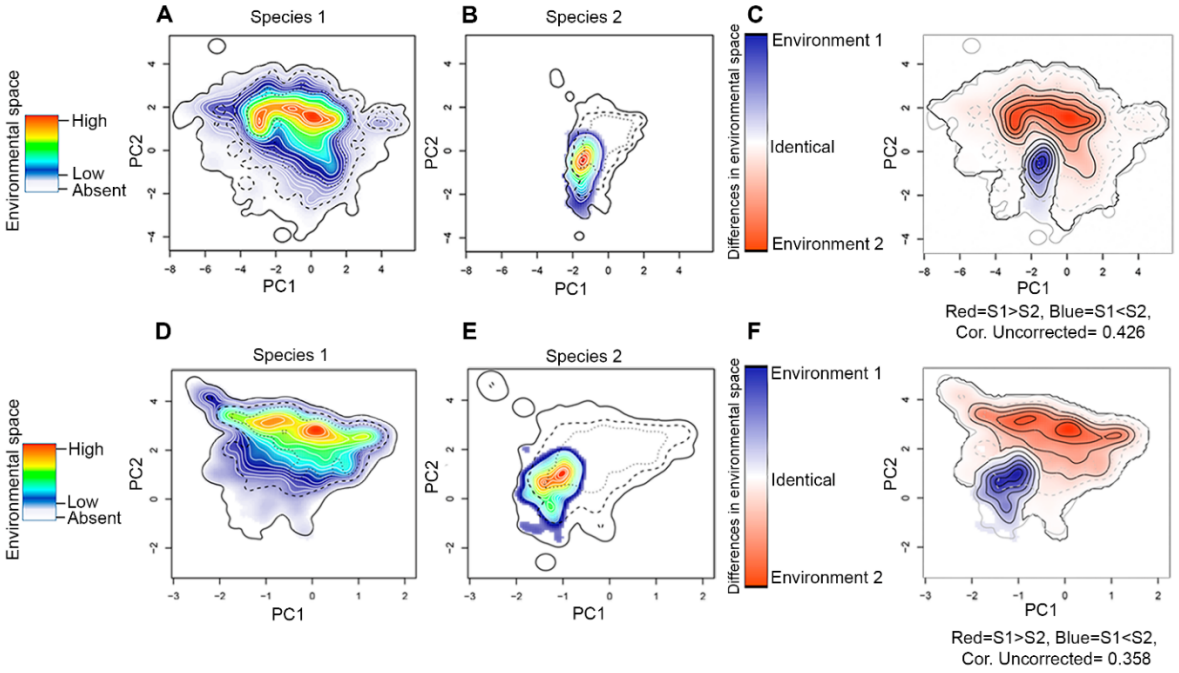

Figure 1.5 Assessment of niche similarity. As an example, we show the Niche Overlap Test (NOT–top row) and Niche Divergence Test (NDT–bottom row) between *H. erato* (species 1) and *H. himera* (species 2), but other comparisons can be found in ZENODO (https://doi.org/10.5281/zenodo.5149294). (A, D) Environmental space of species 1; (B, E) environmental space of species 2; and (C, F) difference in the environmental space (E-space) of two species and Niche E-space Correlation Index (NECI). When NECI was higher than 0.5, we corrected species occupied niches by the frequency of E-space in accessible environments. Significance of NOT and NDT can be found in Table 1.1. Equivalency statistic and niche background statistic for each NOT and NDT can be found in https://doi.org/10.5281/zenodo.5149294.

The foothills of the eastern Andes and the Amazon basin appeared as the regions with highest *Heliconius* species richness, which confirms the findings of a previous study done for the subfamily Heliconiine at a higher scale (50 Km) (28). Interestingly,

both of these regions are known to present unusual concentrations of contact zones and hybrid zones (i.e. suture zones) (77,78), which may explain the richness they exhibit. Also, altitude, isothermality and precipitation were the variables best correlated with this metric. This may be due to the elevational gradient found at the foothills of the eastern Andes, which offers multiple ecological niches thus favoring diversification rates (79–82). Additionally, there are several climate-based hypotheses that seek to explain broad-scale diversity patterns, and water and energy have emerged as crucial influencers of species richness (83). In particular, the water-energy dynamics hypothesis argues that species richness increases in places where liquid water and optimal energy conditions provide the greatest capacity for biotic dynamics (84). The Amazon and foothills of the eastern Andes are regions with near constant hot-warm temperature throughout the year and have a permanent liquid water supply (8,27) thus ensuring an optimal water-energy dynamic. The latter translates into constant availability of plants for butterflies, including host plants for immature and pollen for adults, and continual interactions between individuals, which may be correlated with the high species richness we detected.

Similar to other studies, patterns of phylogenetic diversity were similar (although not identical) to richness (85–88). Interestingly, areas with highest species richness got low phylogenetic diversity (Figure 1.2C, blue grids), which may be a consequence of the recent increase in diversification rate in *Heliconius* (4.5 Ma) and the consequent co-occurrence of multiple young species in the Amazon and foothills of the eastern Andes (28,54). In agreement with this observation, previous research in both animals and plants have found high phylogenetic diversity in the eastern Andes of Colombia, Peru and Ecuador (88–92).

The highest phylogenetic endemism was found in the central eastern Andes of Colombia, and this result is possibly due to the restricted range of the species *Heliconius heurippa* (Figure 1.2D, area H1). However, we cannot rule out this result as an overestimation since the phylogenetic tree that we used (54) considers this taxon as a separate species and not as part of *H. timareta* (as recently hypothesized). If *H. heurippa* had been included within *H. timareta*, which has a wider distribution range, it is likely this result on phylogenetic endemism does not

hold. Additionally, the pacific region of Costa Rica, Panama and Colombia show intermediate values of phylogenetic endemism that resulted from the presence of species that have reduced geographic range and are either long-branch species (e.g., *Heliconius godmani*) or species for which no close relatives are known (e.g., *Heliconius hewitsoni*) (Figure 1.2D, area H2 and H3 respectively). These regions were previously described as highly endemic phylogenetically for plants (93), terrestrial mammals (94), birds and amphibians (95). Interestingly, there were several species that, although are considered as geographically endemic within *Heliconius*, exhibited low values of phylogenetic endemism. However, it is important to acknowledge that phylogenetic endemism is a concept based on linages rather than species, and thus, if an endemic species has a narrow range but it is closely related to a widespread species, its phylogenetic endemism will not necessarily be low (96). An example of this is *Heliconius nattereri,* an endemic species from Brazil's Atlantic Forest that, despite having a narrow distribution, is sister to the widely distributed *Heliconius ethilla* (Figure 1.2D, area H4). Similarly, *Heliconius atthis* is restricted to the Ecuadorian and Peruvian Pacific, but it is sister to the widely distributed *Heliconius hecale* (Figure 1.2D, area H5). In our study we found that high precipitation and near constant hot-warm temperature throughout the year are strongly correlated with phylogenetic endemism, which agrees with studies that point a role for temperature in promoting endemism by reducing extinction rates and increasing population sizes in small areas (94,97,98).

Our environmental niche analysis showed that hybridizing species do not necessarily share the same climatic space despite some of them having largely overlapping geographic distributions. This is the case of *H. ethilla* and *H. numata,* which frequently co-occur throughout their distribution, but there are some regions with an extreme climate such as the Pacific coast of Colombia (a humid jungle) and the Colombian Magdalena valley (which has a marked precipitation gradient, being humid in the north and dry in the south), where *H. ethilla* but not *H. numata* occur (Figure S1.9). This suggests that the former species has a broader climatic tolerance. We also detected differences in the environmental niche between pairs of hybridizing species that rarely overlap geographically, but when they do, they

hybridize. For example, *H. erato* and *H. himera* occupy contrasting environmental niches in Ecuador (99), where *H. himera* lives in dry forests while *H. erato* inhabits wet forests of the Andes (Figure 1.5). Similarly, the hybridizing *H. erato* (*H. e. venus*) and *H. chestertonii* meet in an environmental transition zone between wet and dry forest in the Colombian Andes (100); Figure S1.8).

In summary, we confirmed that, at large scales, the distribution of *Heliconius*, its richness, diversity, and phylogenetic endemism are mainly shaped by a combination of high annual energy (i.e., hot-warm temperature), constant water supply, and an extraordinary topographic complexity. However, species distributions are thought to result from dynamics occurring at multiple spatial scales. Therefore, including microclimate variables and ecological interactions would provide an in-depth understanding of the multiscale drivers of distribution, niche range and phylogenetic processes (76,101). Our study confirms the richness and diversity of areas already identified in other taxa, thus strengthening the importance for their conservation as strategic hotspots of biodiversity.

References

1. Brown JL, Paz A, Reginato M, Renata CA, Assis C, Lyra M, et al. Seeing the forest through many trees: Multi-taxon patterns of phylogenetic diversity in the Atlantic Forest hotspot. Divers Distrib. 2020;26(9):1160–76.
2. Gotelli NJ, Anderson MJ, Arita HT, Chao A, Colwell RK, Connolly SR, et al. Patterns and causes of species richness: A general simulation model for macroecology. Ecol Lett. 2009;12(9):873–86.
3. Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan JF, Kaufman DM, et al. Energy, water, and broad-scale geographic patterns of species richness. Ecology. 2003;84(12):3105–17.
4. Kreft H, Jetz W. Global patterns and determinants of vascular plant diversity. Proc Natl Acad Sci U S A. 2007;104(14):5925–30.

5. Vasconcelos TS, da Silva FR, dos Santos TG, Prado VHM, Provete DB. Biogeographic Patterns of South American Anurans. Biogeographic Patterns of South American Anurans. Springer; 2019. 155 p.
6. Qian H. Environment-richness relationships for mammals, birds, reptiles, and amphibians at global and regional scales. Ecol Res. 2010;25(3):629–37.
7. Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan JF, Kaufman DM, et al. Energy, water, and broad-scale geographic patterns of species richness. Ecology. 2003;84(12):3105–17.
8. Vallejos-Garrido P, Rivera R, Inostroza-Michae O, Rodríguez-Serrano E, Hernández CE. Historical dynamics and current environmental effects explain the spatial distribution of species richness patterns of New World monkeys. PeerJ. 2017;5:2–27.
9. Velazco SJE, Svenning JC, Ribeiro BR, Laureto LMO. On opportunities and threats to conserve the phylogenetic diversity of Neotropical palms. Divers Distrib. 2021;27(3):512–23.
10. Fenker J, Tedeschi LG, Pyron RA, Nogueira C de C. Phylogenetic diversity, habitat loss and conservation in South American pitvipers (Crotalinae: Bothrops and Bothrocophias). Divers Distrib. 2014;20(10):1108–19.
11. Kreft H, Jetz W. Global patterns and determinants of vascular plant diversity. Proc Natl Acad Sci U S A. 2007;104(14):5925–30.
12. Varzinczak LH, Zanata TB, Moura MO, Passos FC. Geographical patterns and current and short-term historical correlates of phylogenetic diversity and endemism for New World primates. J Biogeogr. 2020;47(4):890–902.
13. Rosauer DF, Jetz W. Phylogenetic endemism in terrestrial mammals. Glob Ecol Biogeogr. 2014;24(2):168–79.
14. López-Aguirre C, Hand SJ, Laffan SW, Archer M. Zoogeographical regions and geospatial patterns of phylogenetic diversity and endemism of New World bats. Ecography. 2019;42(6):1188–99.
15. Mullen SP, Savage WK, Wahlberg N, Willmott KR. Rapid diversification and not clade age explains high diversity in neotropical Adelpha butterflies. Proc R Soc B Biol Sci. 2011;278(1713):1777–85.

16. Pearson DL, Carroll SS. Predicting Patterns of Tiger Beetle (Coleoptera : Cicindelidae) Species Richness in Northwestern South America. Stud Neotropical Fauna Environ. 2001;36(2):125–36.

17. Andújar C, Arribas P, Ruiz C, Serrano J, Gómez-Zurita J. Integration of conflict into integrative taxonomy: fitting hybridization in species delimitation ofMesocarabus(Coleoptera: Carabidae). Mol Ecol. 2014;23(17):Andújar, C., Arribas, P., Ruiz, C., Serrano, J., .

18. Arteaga MC, Mccormack JE, Eguiarte LE, Medell RA. Genetic admixture in multidimensional enviromental space : asymetrical niche similarity promotes gene flow in Armadillos (Dasypus Novemcinctus). Evolution. 2011;65(9):2470–80.

19. Ortego J, Gugger PF, Riordan EC, Sork VL. Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. J Biogeogr. 2014;41(10):1895–908.

20. Swenson NG, Fair JM, Heikoop J. Water stress and hybridization between Quercus gambelii and Quercus grisea. West North Am Nat. 2008;68:498–507.

21. Jiggins C. The ecology and evolution of *Heliconius* butterflies. Oxford University Press; 2017. 288 p.

22. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 2013;23(11):1817–28.

23. Kronforst M.R., Hansen M.E., Crawford N.G., Gallant J.R., Zhang, W., Kulathinal R.J., Kapan D.D., Mullen S.P. Hybridization reveals the evolving genomic architecture of speciation. Cell Rep. 2013;14(5):666–77.

24. Mallet J, Beltrán M, Neukirchen W, Linares M. Natural hybridization in heliconiine butterflies: The species boundary as a continuum. BMC Evol Biol. 2007;7:1–16.

25. Mallet J, McMillan WO, Jiggins CD. Mimicry and Warning Color at the Boundary between Races and Species. In: Endless forms: species and speciation. New York: Oxford Univ. Press; 1998. p. 390–403.

26. Mallet J, Barton NH, Lamas G, Santisteban J, Muedas M, Eeley H. Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics. 1990;124:921–36.

27. Rosser N, Dasmahapatra KK, Mallet J. Stable *Heliconius* butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. Evolution. 2014;68(12):3470–84.

28. Rosser N, Phillimore A, Huertas B, Willmott K, Mallet J. Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. Biol J Linn Soc. 2012;105(March):479–97.

29. Lamas G, Jiggins CD. Taxonomic list. In: The ecology & evolution of *Heliconius* butterflies. 2017. p. 214–44.

30. Wieczorek J, Guo Q, Hijmans RJ. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. Int J Geogr Inf Sci. 2004;18(8):745–67.

31. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Climatologies at high resolution for the earth's land surface areas. Sci Data. 2017 Sep 5;4:170122.

32. Jarvis A, Reuter HI, Nelson A, Guevara E. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database [Internet]. 2008. Available from: http://srtm.csi.cgiar.org

33. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.r-project.org/

34. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography. 2013;36(1):27–46.

35. Heiberger MRM. Package ' HH '. 2020.

36. Kubota Y, Shiono T, Kusumoto B. Role of climate and geohistorical factors in driving plant richness patterns and endemicity on the east Asian continental islands. Ecography. 2015;38(6):639–48.

37. Assis J. R Pipelines to reduce the spatial autocorrelation in Species Distribution Models. [Internet]. theMarineDataScientist. 2020 [cited 2021 Apr 20]. Available from: https://github.com/jorgeassis/spatialAutocorrelation
38. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: How, where and how many? Methods Ecol Evol. 2012;3(2):327–38.
39. Lake TA, Briscoe Runquist RD, Moeller DA. Predicting range expansion of invasive species: Pitfalls and best practices for obtaining biologically realistic projections. Divers Distrib. 2020;26(12):1767–79.
40. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecol Appl. 2009;19(1):181–97.
41. Soberón J, Nakamura M. Niches and distributional areas: Concepts, methods, and assumptions. Proc Natl Acad Sci U S A. 2009;106(SUPPL. 2):19644–50.
42. Schmitt S, Pouteau R, Justeau D, de Boissieu F, Birnbaum P. ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. Methods Ecol Evol. 2017;8(12):1795–803.
43. McCullagh P, Nelder JA. Generalized linear models. London, UK; 1989.
44. Friedman JH, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat. 2000;28:337–407.
45. Phillips, Steven Anderson, Robert Schapire R. Maximum entropy modeling of species geographic distributions. Ecol Model. 2006;190:231–59.
46. Hastie TJ, Tibshirani RJ. Generalized Additive Models. London, Reino Unido: Chapman and Hall/CRC; 1990. 352 p.
47. Smith AB, Santos MJ. Testing the ability of species distribution models to infer variable importance. Ecography. 2020;43(12):1801–13.
48. Schmitt S, Pouteau R, Justeau D, de Boissieu F, Birnbaum P. ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. Methods Ecol Evol. 2017;8(12):1795–803.

49. Daru BH, Karunarathne P, Schliep K. phyloregion: R package for biogeographical regionalization and macroecology. Methods Ecol Evol. 2020;11(11):1483–91.
50. Davis Rabosky AR, Cox CL, Rabosky DL, Title PO, Holmes IA, Feldman A, et al. Coral snakes predict the evolution of mimicry across New World snakes. Nat Commun. 2016;7(May):1–9.
51. Paz A, Brown JL, Cordeiro CLO, Aguirre-Santoro J, Assis C, Amaro RC, et al. Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic Forest. J Biogeogr. 2021;(March).
52. Meyer L, Diniz-Filho JAF, Lohmann LG. A comparison of hull methods for estimating species ranges and richness maps. Plant Ecol Divers. 2017;10(5–6):389–401.
53. Daru BH, Karunarathne P, Schliep K. phyloregion: R package for biogeographical regionalization and macroecology. Methods Ecol Evol. 2020;11(11):1483–91.
54. Kozak KM, Wahlberg N, Neild AF, Dasmahapatra KK, Mallet J, Jiggins CD. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst Biol. 2015;64(3):505–24.
55. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61(1):1–10.
56. Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG. Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. Mol Ecol. 2009;18(19):4061–72.
57. Paz A, Brown JL, Cordeiro CLO, Aguirre-Santoro J, Assis C, Amaro RC, et al. Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic Forest. J Biogeogr. 2021;(March).
58. Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(December):18–22.
59. Venables WN, Ripley BD. Modern applied statistics with S. Vol. 53. Springer; 2002. 481 p.

60. Karatzoglou A, Hornik K, Smola A, Zeileis A. kernlab - An S4 package for kernel methods in R. J Stat Softw. 2004;11:1–20.
61. Kuhn M. caret Package. J Stat Softw. 2008;28(5):1–26.
62. Brown JL, Carnaval AC. A tale of two niches: Methods, concepts, and evolution. Front Biogeogr. 2019;11(4):1–27.
63. Soberón J, Nakamura M. Niches and distributional areas: Concepts, methods, and assumptions. Proc Natl Acad Sci U S A. 2009;106(SUPPL. 2):19644–50.
64. Rödder D, Engler JO. Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. Glob Ecol Biogeogr. 2011;20(6):915–27.
65. O'Donnell MS, Ignizio DA. Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States. US Geol Surv Data Ser 691. 2012;1–17.
66. Chattopadhyay B, Garg KM, Ray R, Rheindt FE. Fluctuating fortunes: Genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity. Proc R Soc B Biol Sci. 2019;286(1911):1–10.
67. Heya HM, Khamis FM, Onyambu GK, Akutse KS, Mohamed SA, Kimathi EK, et al. Characterization and risk assessment of the invasive papaya mealybug, Paracoccus marginatus, in Kenya under changing climate. J Appl Entomol. 2020;144(6):1–17.
68. Simó M, Guerrero JC, Giuliani L, Castellano I, Acosta LE. A predictive modeling approach to test distributional uniformity of Uruguayan harvestmen (Arachnida: Opiliones). Zool Stud. 2014;53(1):1–13.
69. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci Data. 2018;5:1–12.
70. Cormont A, Malinowska AH, Kostenko O, Radchuk V, Hemerik L, WallisDeVries MF, et al. Effect of local weather on butterfly flight behaviour, movement, and colonization: Significance for dispersal under climate change. Biodivers Conserv. 2011;20(3):483–503.

71. Amundrud SL, Videla M, Srivastava DS. Dispersal barriers and climate determine the geographic distribution of the helicopter damselfly Mecistogaster modesta. Freshw Biol. 2018;63(2):214–23.
72. Atauchi PJ, Peterson T, Flanagan J. Species distribution models for Peruvian Plantcutter improve with consideration of biotic interactions. J Avian Biol. 2017;49(3).
73. de Oliveira da Conceição E, Mantovano T, de Campos R, Rangel TF, Martens K, Bailly D, et al. Mapping the observed and modelled intracontinental distribution of non-marine ostracods from South America. Hydrobiologia. 2020;847(7):1663–87.
74. Schivo F, Bauni V, Krug P, Quintana RD. Distribution and richness of amphibians under different climate change scenarios in a subtropical region of South America. Appl Geogr. 2019;103(February 2018):70–89.
75. Montejo-Kovacevich G, Smith JE, Meier JI, Bacquet CN, Whiltshire-Romero E, Nadeau NJ, et al. Altitude and life-history shape the evolution of *Heliconius* wings. Evolution. 2019;73(12):2436–50.
76. Montejo-Kovacevich G, Martin SH, Meier JI, Bacquet CN, Monllor M, Jiggins CD, et al. Microclimate buffering and thermal tolerance across elevations in a tropical butterfly. J Exp Biol. 2020;223(8):jeb220426.
77. Dasmahapatra KK, Lamas G, Simpson F, Mallet J. The anatomy of a 'suture zone' in Amazonian butterflies: A coalescent-based test for vicariant geographic divergence and speciation. Mol Ecol. 2010;19(19):4283–301.
78. Rosser N, Shirai LT, Dasmahapatra KK, Mallet J, Freitas AVL. The Amazon river is a suture zone for a polyphyletic group of co-mimetic heliconiine butterflies. Ecography. 2021;44:177–87.
79. Davies RG, Orme CDL, Storch D, Olson VA, Thomas GH, Ross SG, et al. Topography, energy and the global distribution of bird species richness. Proc R Soc B Biol Sci. 2007;274(1614):1189–97.
80. Jetz W, Rahbek C. Geographic range size and determinants of avian species richness. Science. 2002;297(5586):1548–51.
81. Keppel G, Gillespie TW, Ormerod P, Fricker GA. Habitat diversity predicts orchid diversity in the tropical south-west Pacific. J Biogeogr. 2016;43(12):2332–42.

82. Rahbek C, Graves GR. Multiscale assessment of patterns of avian species richness. Proc Natl Acad Sci U S A. 2001;98(8):4534–9.
83. Silva-Flores R, Pérez-Verdín G, Wehenkel C. Patterns of tree species diversity in relation to climatic factors on the Sierra Madre Occidental, Mexico. PLoS ONE. 2014;9(8).
84. Svenning JC, Borchsenius F, Bjorholm S, Balslev H. High tropical net diversification drives the New World latitudinal gradient in palm (Arecaceae) species richness. J Biogeogr. 2008;35(3):394–406.
85. Davies Jonathan T, Buckley LB. Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. Philos Trans R Soc B Biol Sci. 2011;366(1576):2414–25.
86. Guedes TB, Sawaya RJ, Zizka A, Laffan S, Faurby S, Pyron RA, et al. Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. Glob Ecol Biogeogr. 2018;27(1):14–21.
87. Mendoza AM, Arita HT. Priority setting by sites and by species using rarity, richness and phylogenetic diversity: The case of neotropical glassfrogs (Anura: Centrolenidae). Biodivers Conserv. 2014;23(4):909–26.
88. Fenker J, Tedeschi LG, Pyron RA, Nogueira C de C. Phylogenetic diversity, habitat loss and conservation in South American pitvipers (Crotalinae: Bothrops and Bothrocophias). Divers Distrib. 2014;20(10):1108–19.
89. Arango A, Villalobos F, Prieto-Torres DA, Guevara R. The phylogenetic diversity and structure of the seasonally dry forests in the Neotropics. J Biogeogr. 2021;48(1):176–86.
90. Guedes TB, Sawaya RJ, Zizka A, Laffan S, Faurby S, Pyron RA, et al. Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. Glob Ecol Biogeogr. 2018;27(1):14–21.
91. Mendoza AM, Arita HT. Priority setting by sites and by species using rarity, richness and phylogenetic diversity: The case of neotropical glassfrogs (Anura: Centrolenidae). Biodivers Conserv. 2014;23(4):909–26.

92. Velazco SJE, Svenning JC, Ribeiro BR, Laureto LMO. On opportunities and threats to conserve the phylogenetic diversity of Neotropical palms. Divers Distrib. 2021;27(3):512–23.
93. Sandel B, Weigelt P, Kreft H, Keppel G, van der Sande MT, Levin S, et al. Current climate, isolation and history drive global patterns of tree phylogenetic endemism. Glob Ecol Biogeogr. 2020;29(1):4–15.
94. Rosauer DF, Jetz W. Phylogenetic endemism in terrestrial mammals. Glob Ecol Biogeogr. 2014;24(2):168–79.
95. Daru BH, Farooq H, Antonelli A, Faurby S. Endemism patterns are scale dependent. Nat Commun. 2020;11(1):1–11.
96. Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG. Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. Mol Ecol. 2009;18(19):4061–72.
97. Jetz W, Rahbek C, Colwell RK. The coincidence of rarity and richness and the potential signature of history in centres of endemism. Ecol Lett. 2004;7(12):1180–91.
98. Varzinczak LH, Zanata TB, Moura MO, Passos FC. Geographical patterns and current and short-term historical correlates of phylogenetic diversity and endemism for New World primates. J Biogeogr. 2020;47(4):890–902.
99. Jiggins C, Mcmillan WO, Mallet J. Host plant adaptation has not played a role in the recent speciation of *Heliconius himera* and *Heliconius erato*. Ecol Entomol. 1997;22(3):361–5.
100. Muñoz AG, Salazar C, Castaño J, Jiggins CD, Linares M. Multiple sources of reproductive isolation in a bimodal butterfly hybrid zone. J Evol Biol. 2010;23(6):1312–20.
101. Paz A, Guarnizo CE. Environmental ranges estimated from species distribution models are not good predictors of lizard and frog physiological tolerances. Evol Ecol. 2020;34(1):89–99.

CHAPTER 2.

Three potential sex chromosome – autosome fusions in *Heliconius* butterflies

Introduction

Sex chromosome-autosome (SA) fusions contribute to the turnover of sex-determination loci or the evolution of neo-sex chromosomes (1–3), but it remains unclear what promotes them. Sexually antagonistic selection, the sheltering of deleterious mutations, direct selection, genetic drift or meiotic drive have all been suggested as possible drivers of SA fusions (4–7). Sexually antagonistic selection is thought to favour the fusion of sex chromosomes with autosomes harbouring genes under sexually antagonistic selection (8). There is limited evidence for this theory e.g. in sticklebacks (1) and warblers (9). An alternative hypothesis is deleterious mutation sheltering, where SA fusions are favoured because they prevent the expression of recessive deleterious alleles in the heterogametic sex (4). SA fusions can also become fixed due to meiotic drive (or holocentric drive in holocentric organisms (10), such as female meiotic drive elements on W/X-A fusions that preferentially end up in the egg instead of the polar bodies (11).

As with other chromosomal rearrangements, SA fusions can reduce recombination and potentially strengthen reproductive isolation if they bring together barrier loci into regions with reduced recombination (12,13). For instance, in the Japanese threespine stickleback *Gasterosteus aculeatus,* an SA fusion resulted in a neo-X chromosome that linked loci underlying behavioural isolation traits and hybrid sterility (5). SA fusions may also facilitate adaptation, such as the SA fusion in *Cydia pomonella* (Tortricidae), which apparently linked two insecticide-resistance genes and genes involved in detoxifying plant metabolites (14).

Genomic studies have revealed that SA fusions have occurred multiple times across vertebrates (7,15,16), and in some invertebrates such as mygalomorph spiders (17), *Drosophila* (18) and true bugs of the genus *Dysdercus* (19). In Lepidoptera, examples of SA fusions include *Danaus* (20,21) and *Leptidea* butterflies (3,22), among others (14,23–25). Compared to taxa with a single centromere per chromosome, the holocentric chromosomes of Lepidoptera may facilitate fusions, as they are less likely to cause segregation problems during cytokinesis and thus reduce hybrid fitness (26). Nonetheless, butterflies and moths have remarkably constrained chromosome evolution (27), with most species having a ZW or Z0 sex determination system and a haploid chromosome number ranging between 28 and 32 except for a few groups that have experienced extensive fission and fusion events (28). One of such cases are the *Heliconius* butterflies, which mostly have 21 chromosomes due to ancestral fusions (29). Only a few species in the genus differ from this ancestral chromosome number, especially species in the *sara/sapho* clade, with some having up to 60 chromosomes (30). This clade comprises 12 species (31), and differs from other *Heliconius* by showing an inability to synthesise cyanogens, leading to reliance on toxins sequestered from larval host plants (32,33). A subclade of seven species within the *sara/sapho* clade shows particularly high diversification rates (34) and high number of chromosomes (30).

We generated whole genome resequencing data for 114 individuals from these seven species in the *sara/sapho* clade to: (i) investigate genome-wide phylogenetic relationships, (ii) determine the degree of genomic differentiation between species and subspecies, and (iii) explore the impact of chromosomal rearrangements. Despite a high number of chromosomes in this clade, suggesting many chromosomal fissions, we find at least three SA fusions across five species, making them a prime study system for the evolution of neo-sex chromosomes.

Materials and Methods

Genome assembly of Heliconius sara

We used a female of *H. sara* (SAMEA5394385; Table S2.1) to generate an assembled genome for this species. We assembled the genome by combining PacBio, HiC and 10X data, all generated by the Tree of Life Programme and the Wellcome Sanger Institute (https://www.sanger.ac.uk/programme/tree-of-life/). The libraries were produced following the manufacturer's guidelines. We sequenced four SMRT (single-molecule real-time) cells with the PacBio Sequel II system. Dovetail Hi-C data was sequenced on a HiSeq X Ten platform. A 10X Genomics Chromium linked-read sequencing library was sequenced with 150-bp paired-end reads on four lanes on an Illumina HiSeq X Ten platform.

An Initial contig assembly was generated from the PacBio CLR data using wtdbg2 v2.2 (35) . The PacBio data was then used to polish the contigs using Arrow (https://github.com/PacificBiosciences/GenomicConsensus). The 10X Chromium linked-reads were then mapped to the assembly using Longranger v2.2. (10X Genomics), variants called using freebayes v1.1.0-3-g961e5f3 (36) . Next, the assembly was polished using BCFtools consensus v1.9 (37) by applying homozygous non-reference calls as edits. The 10X linked-reads were then used to scaffold contigs using Scaff10X v2.3 (https://github.com/wtsi-hpag/Scaff10X). A round of manual curation was performed on these polished scaffolds using gEVAL (38) . Lastly, Dovetail Genomics Hi-C data was used to scaffold the assembly further using SALSA v2.2 (39), followed by another round of manual curation with gEVAL (38). The chromosome-scale scaffolds were named by synteny to the *Heliconius melpomene melpomene* assembly Hmel2.5 in LepBase. We assessed the genome contiguity with gnx-tools (https://github.com/mh11/gnx-tools/blob/master/README) and genome completeness with BUSCO v5.1.2 (40) using the Lepidoptera gene set.

Sample collection for genome resequencing

We collected 114 individuals of *Heliconius* from 7 species and 18 subspecies in the *sara/sapho* clade across their distribution range: 48 *H. sara*, 2 *H. leucadia*, 21 *H. antiochus*, 13 *H. sapho*, 3 *H. hewitsoni*, 17 *H. eleuchia* and 10 *H. congener* (Table

S2.1). The body of each was preserved in NaCl-saturated DMSO solution and stored at -80°C; wings were kept for phenotype reference.

Whole-genome resequencing and genotype calling

Genomic DNA was extracted from thoracic tissue using a DNeasy Blood and Tissue Kit (Qiagen). Library preparation and whole-genome Illumina resequencing (PE reads) was carried out on Illumina's HiSeq X system by Novogene (Beijing, China), with 30X coverage per individual. We also downloaded two samples of *H. charithonia* (SRR4032025 – SRR4032026) from SRA (https://www.ncbi.nlm.nih.gov/sra) to include them as outgroups in phylogenetic analyses. The *H. sara* genome (HelSar1) was used as a reference to map the reads of each individual using BWA mem v0.7.12 (41) with default parameters. We then used samtools v1.12 to sort and index the alignment files (42). PCR-duplicate reads were identified and removed using Picard tools v2.9.2 (43), and variant calling was conducted with Haplotype Caller (GATK, v3.7.0) in BP-resolution mode (44). Then, samples were jointly genotyped using GATK's GenotypeGVCFs (44). We used vcftools v0.1.14 (45) and the final VCF to calculate: (i) mean depth per individual and site, (ii) quality per site, (iii) the proportion of missing data per individual and (iv) the proportion of missing data per site, and (v) percentage mapping per individual. Based on these results, we kept sites with quality value (--minQ) ≥ 30 and less than 5% missing data. We also excluded sites with a depth below 5 and mean depth per individual more than 1.5 times the mean to exclude paralogous regions (46). For this, we used the custom script *removeTooHighDepthSites.sh* from ((https://github.com/joanam/VictoriaRegionSuperflock/BashPipelines). We removed sites with excess heterozygosity using the vcftools option *--hardy* and a p-value cut-off of <1e-5. Finally, we checked our samples for contamination or other issues generating erroneous heterozygotes using the python script *checkHetsIndvVCF.sh* by David Marques, and modified by Jessica Rick (https://github.com/jessicarick/lates-popgen/blob/master/scripts/).

Phylogenetic analysis

We generated a whole-genome ML tree using a vcf containing all sites as input in RAXML v8.2.9 (47), with the GTRGAMMA model and 100 bootstrap replicates. We applied the same procedure to obtain ML trees for each chromosome. In all cases, *H. charithonia* was set as outgroup.

Population structure and shared ancestry

We performed a principal component analysis (PCA) to study the genetic structure of populations. We filtered out monomorphic or multiallelic sites, and sites with minor allele frequency (MAF) smaller than 0.1 with vcftools (45). To reduce the linkage disequilibrium effect, we used the python script *ldPruning.sh* from (https://github.com/joanam/scripts), which removes sites with r^2>0.2 in windows of 50 Kbp sliding by 10 Kbp. This resulted in a vcf file with 3,685,916 high-quality SNPs sites. We conducted the PCA using Plink v2.0 with default parameters (48,49).

We analysed patterns of population structure using the R package *lostruct* (50) on chromosomes revealing sex-clustering phylogenies. We performed this analysis between pairs of species (*H. eleuchia* - *H. congener* and *H. sapho* - *H. hewitsoni*), and between subspecies of *H. antiochus*. We first divided the chromosomes into non-overlapping windows of 100 SNPs and then performed a PCA for each window to identify regions that might indicate variation in the chromosome structure. We next used a two-dimension space multidimensional scaling (MDS) analysis to illustrate how dissimilar the patterns of relatedness are between each pair of windows. We then manually selected sets of windows to calculate PCAs and test which regions do not group samples by species. Finally, we used R (51) to visualise the PCA results and plot the first two MDS coordinates against the midpoint of each window to visualise variation across the genome.

The genetic ancestry of each sample was estimated using ADMIXTURE v. 1.3.0 (52). Input data were prepared using the same procedure for the genome wide

PCA analysis. We used K = 1-6 clusters and the cross-validation method to estimate the best K value. All other parameters were set to default.

Patterns of genetic differentiation

We calculated F_{ST} and D_{XY} by pairs of sister species and by subspecies along chromosomes in non-overlapping 50 Kbp windows. Because *H. antiochus* did not have a sister species, we calculated these statistics between Andean and Amazonian subspecies. Windows that contained less than 2,500 high-quality genotyped variable sites were rejected. We used a dataset including SNPs and monomorphic sites and the popgenWindows.py script from (https://github.com/simonhmartin/genomics_general).

Patterns of heterozygosity and mean depth by chromosome

To study chromosomes with F_{ST} and D_{XY} patterns different from the genome average, we used the options --het and --depth of vcftools v. 0.1.14 (45), and calculated heterozygosity and mean depth per chromosome for each individual of each species. We also calculated these statistics in 50 Kbp non-overlapping sliding windows along the 'outlier' chromosomes identified. On these, we calculated π specifying each individual as its own population, so π became a measure of heterozygosity. This was done with the Python script *popgenWindows.py* from (https://github.com/simonhmartin/genomics_general). We then averaged these values across all individuals of the same sex and species. For depth, we first generated a file containing the mean depth per site averaged across all individuals of the same sex and species using the --site-mean-depth option of vcftools v. 0.1.14 (45). We then used the R package windowscanr v. 0.1 from (https://github.com/tavareshugo/WindowScanR) to calculate the mean of the mean depth per species, per sex, and per window. The few individuals of *H. hewitsoni* and *H. leucadia* were not included in the sliding windows analysis. Statistical tests were applied to assess significant differences in heterozygosity and mean depth between

sexes and between chromosomes. As the data were not normally distributed, we performed a Wilcoxon signed rank test to compare sexes. To assess differences between chromosomes, we applied a Kruskal-Wallis test and a *post hoc* test (pairwise Wilcoxon test for Kruskal-Wallis).

Results

Genome assembly

The final assembled genome was 349 Mb (with 46,300 bp of Ns) in 399 scaffolds, with scaffold size increasing from 2,007 bp to 22.4 Mb, and with a N50 of 17.8 Mb which is better than the currently best *Heliconius* reference genomes (*H. melpomene*: 14.3 Mb, *H. erato demophoon*: 10.7 Mb). Hi-C assigned 36 scaffolds to 20 chromosomes and the Z chromosome, whilst the remaining 363 scaffolds were unassigned. The BUSCO completeness is 98.4%, of which 98.2% are single-copy BUSCOs.

Whole-genome resequencing dataset

A total of 114 individuals were successfully sequenced and 112 were included in downstream analyses. The average mapping percentage to the *H. sara* genome was 95.56% (range: 77.38% - 99.14%; Table S2.1). Two individuals with a particularly high proportion of missing data (19.5% and 39.5% respectively) and low mean coverage (7.4X and 9.1X, respectively) were excluded from further analyses (Table S2.1 and Figure S2.1). In contrast, *H. sara* and *H. leucadia* had ≤10% missing data per individual, much less than any other species (Figure S2.1) and likely explained by close similarity to the reference genome. The mean coverage of all other samples ranged between 10X and 27X (Table S2.1 and Figure S2.1).

Phylogenetic relationships

The ML phylogenetic tree obtained with 183,282,470 sites separated individuals into groups consistent with both PCA and ADMIXTURE analyses, and revealed two main clades: (i) the *sara* clade and (ii) the *sapho* clade (Figure 2.1). The *sara* clade is composed of two species, namely *H. sara* and *H. leucadia*, where *H. sara* is subdivided into an Andean subgroup (*H. s. magdalena, H. s. sprucei,* and *H. s. elektra*) and an Amazonian subgroup (*H. s. sara*). The *sapho* clade is split into two well-resolved lineages (*H. antiochus* and a clade composed of two monophyletic groups: *H. eleuchia/H. congener* and *H. sapho/H. hewitsoni*). In this case, *H. antiochus* appeared as a monophyletic group split into an Andean clade (*H. a. aranea* and *H. a. araneides*), and an Amazonian clade (*H. a. antiochus* and *H. a. salvini*; Figure 2.1). *H. antiochus* nested into the *sapho* clade whereas *H. hewitsoni* was found as sister to *H. sapho* thus resolving the previously undetermined position of these species (34).

We found strong phylogenetic incongruence across chromosomes. For instance, the whole-genome topology was only recovered in eight chromosomes (Chr1, Chr2, Chr10, Chr11, Chr13, Chr15, Chr18, and Chr19; Figures 2.2A and S2.2-S2.22), while nine chromosomes showed *H. congener* varying its position in the phylogeny and appearing as sister either to *H. e. eleuchia* or to a clade composed of *H. e. eleusinus* + *H. e. primularis* (Chr3, Chr4, Chr6, Chr8, Chr9, Chr12, Chr17, Chr20, and Chr21; Figures 2.2A and S2.2-S2.22). Similarly, the position of *H. hewitsoni* in relation to *H. sapho* varied in eight chromosomes (Chr5, Chr6, Chr7, Chr14, Chr16, Chr17, Chr20, Chr21; Figures 2.2A and S2.2- S2.22).

Interestingly, we found sex-specific clustering on Chr4, Chr9 and Chr14 (Figure 2.2B). In Chr4, all species in the *sapho* clade showed females and males clustering separately, while males of *H. congener* and *H. eleuchia* formed a clade apart from their conspecific females (Figure 2.2B). In the case of *H. antiochus*, *H. sapho*, *H. hewitsoni*, females do not group by sex among themselves because, like *H. congener* and *H. eleuchia*, they are heterozygous for sex-specific haplotypes. However, in these species, the non-female specific copy of Chr4 is highly divergent

between them. This sex clustering pattern was also observed on Chr9 for *H. sapho* and *H. hewitsoni* (Figure 2.2B). In addition, males and females of *H. eleuchia* and *H. congener* also formed a clade in Chr14 (Figure 2.2B).

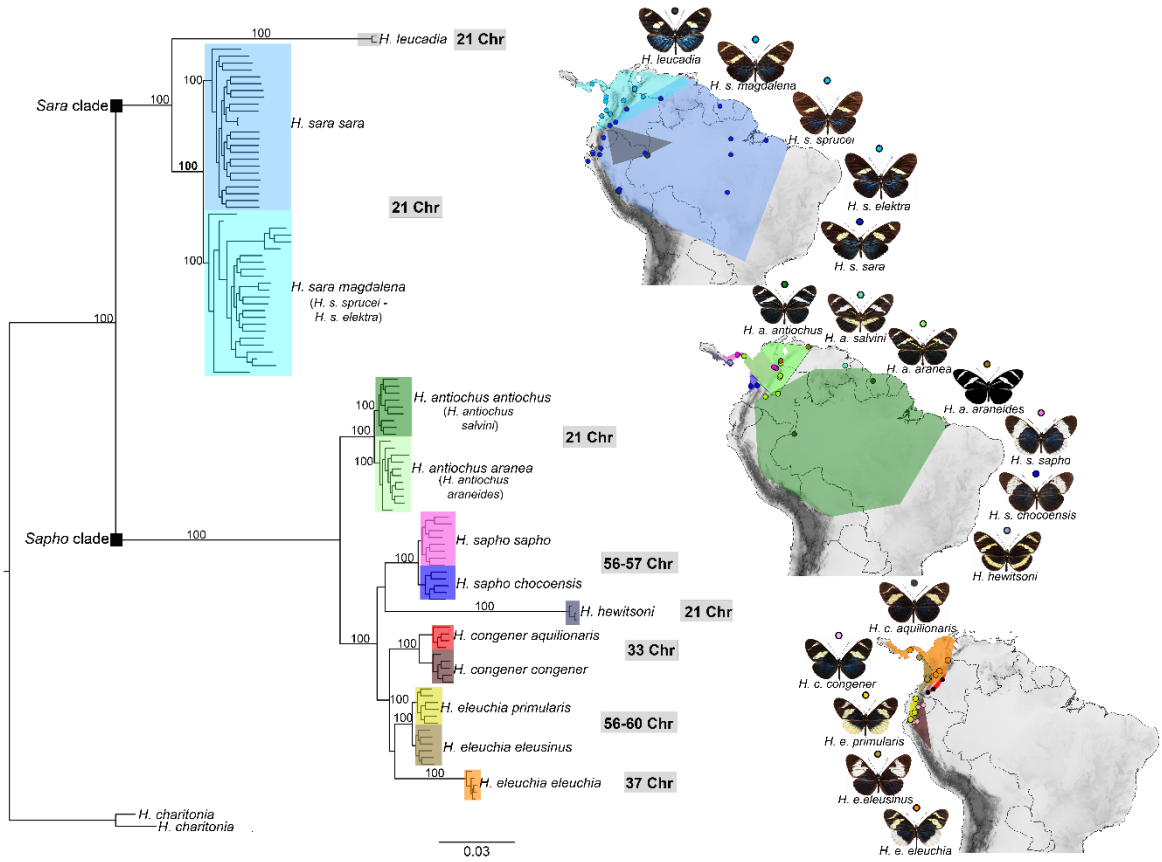

Figure 2.1 Phylogeny and distribution of the *sara/sapho* clade. Genome-wide ML phylogeny, highlighting the position of the *sara* and the *sapho* clade, with bootstrap support values on the branches. The distribution of all 17 subspecies in the clade from (53) is shown, where dots represent sampling locations for the present study. Each of the 17 subspecies is represented by a unique colour, that is the same in the phylogeny and the distribution. The chromosome number of each species is shown in gray rectangles. These values were taken from (54).

Population structure and shared ancestry

The PCA grouped samples into three main groups: (i) *H. sara* and *H. leucadia* (hereafter sara clade), (ii) *H. antiochus*, and (iii) *H. eleuchia, H. congener, H. sapho* and *H. hewitsoni* (hereafter sapho clade). The first two principal components

explained 60% (PC1) and 10% (PC2) of the total variance. PC1 separated the *sara* clade and *H. charithonia* (outgroup) from *H. antiochus* and the *sapho* clade, whereas PC2 separated *H. antiochus* from the *sapho* clade. *H. sapho* was closer to *H. hewitsoni*, whereas *H. eleuchia* was closer to *H. congener* (Figure S2.23). Results from ADMIXTURE were largely consistent with the PCA recovering *H. antiochus*, *H. eleuchia*, *H. sapho* and *H. congener* as independent groups. We also identified the Andean subspecies of *H. sara* (*H. s. magdalena*, *H. s. sprucei*, and *H. s. elektra*) and the Amazonian subspecies of *H. sara* (*H. s. sara*) as separate groups. *H. hewitsoni* and *H. leucadia* could not be correctly assigned as a separate group due to the low number of samples. ADMIXTURE analysis did not reveal any evidence of recent introgression among species (Figure S2.24).

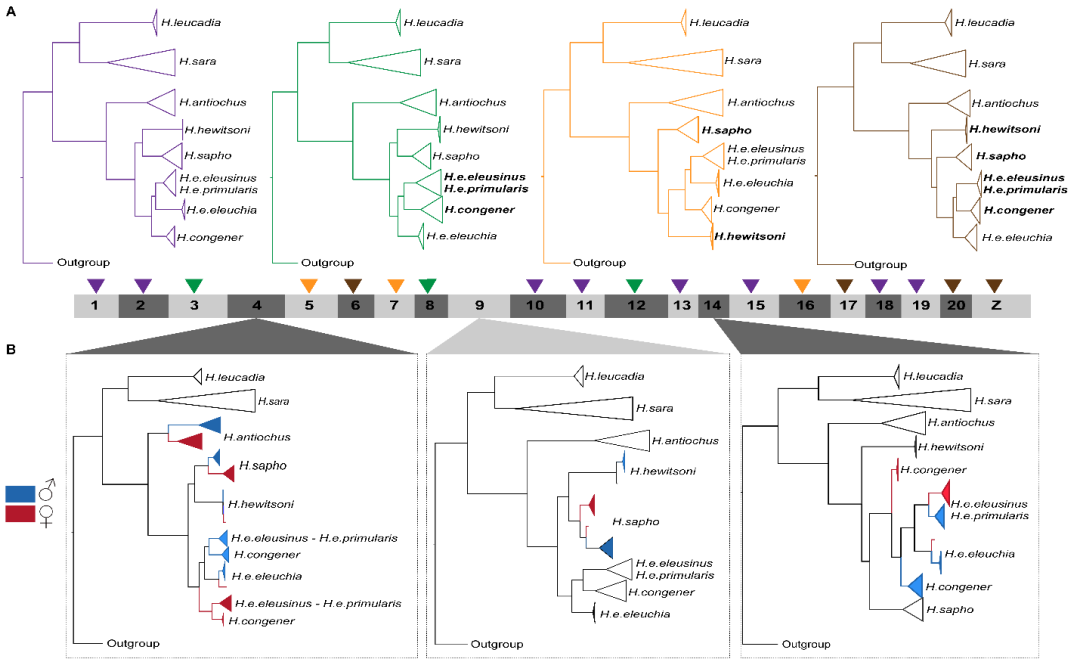

Figure 2.2 ML phylogenies inferred genome-wide and per chromosome. A) Topologies found across the genome. Purple: genome-wide topology. Green: *H. congener* within *H. eleuchia*. Orange: *H. hewitsoni* as sister to *H. congener* + *H. eleuchia*. Brown: *H. hewitsoni* as sister to *H. congener* + *H. eleuchia* + *H. sapho*. Chromosomes are shown in the bottom, with coloured triangles indicating the topology revealed by each of them. B) Topologies showing sex-specific grouping within some species, which is indicative of SA fusions in Chr4, Chr9 and Chr14. In these species, females are coloured in red and males in blue.

The local PCA confirmed the sex specific patterns on Chr4, Chr 9 and Chr14 (Figures 2.3 and S2.25-S2.29). The MDS analysis summarising the PCA results of each 100 SNP window on Chr4 in *H. eleuchia*, *H. congener*, *H. sapho* and *H. hewitsoni* revealed two clusters, one of windows grouping the individuals by species, and the second grouping the individuals by sex (Figures 2.3A-2.3B and 2.3D-2.3E). In *H. antiochus*, all windows in Chr4 cluster individuals by sex (Figure S2.27). Chr14 behaves like Chr4 but only in *H. eleuchia* and *H. congener,* where sex-grouping is observed across the entire chromosome except for the ends, where grouping is by species (Figure S2.28). Finally, Chr9 shows a pattern where PC1 of each window groups *H. sapho* and *H. hewitsoni* individuals by species and PC2 groups them by sex (Figure S2.29). Strikingly, in all species, sex-grouping windows are scattered along Chr4, Chr9 and Chr14 (Figures 2.3C, 2.3F and S2.28C).

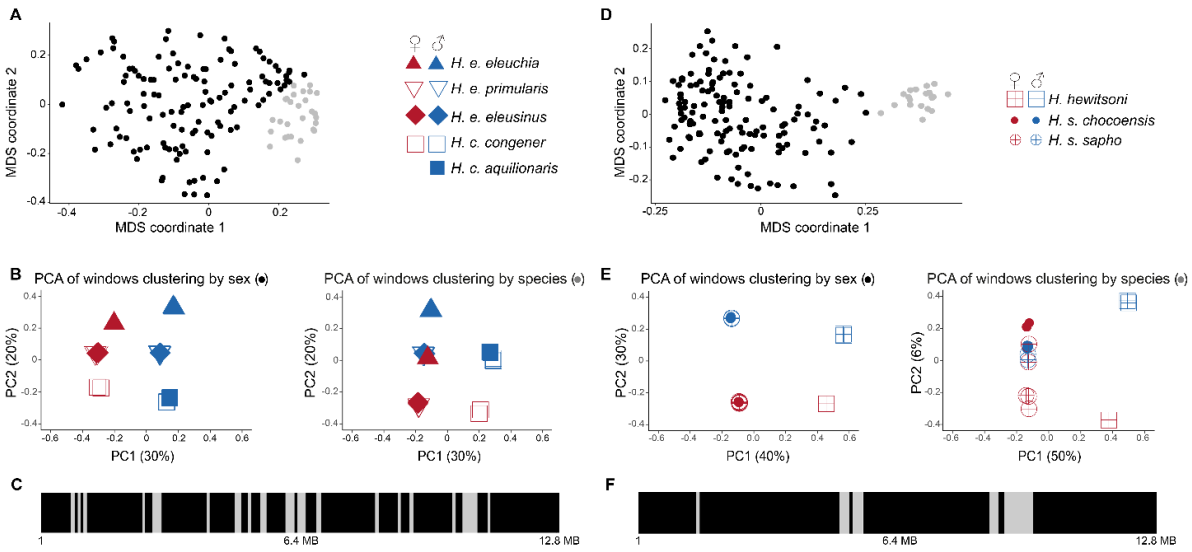

Figure 2.3 Local PCA in 100 SNP windows on Chr4 reveals grouping by sex across the entire chromosome. A) two-dimensional MDS plot of the distribution of 100 SNP windows along Chr4 for the *H. congener/H. eleuchia* clade, B) PCA plots for the black and grey windows in A). C) Actual distribution for the black and grey windows in A) along Chr4. D) two-dimensional MDS plot of the distribution of 100 SNP windows along Chr4 for the *H. sapho/H. hewitsoni* clade. E) PCA plots for the black and grey windows in D). F) Actual distribution for the black and grey windows in D) along Chr4. Each point in A) and D) represents one window. Black: windows that cluster by sex in the local PCA. Grey: windows that cluster by species in the local PCA. Each individual in B) and E) is represented by a point with a species-specific symbol and colour according to its sex. The female symbol for

H. congener aquilionaris is missing because there are no female specimens of that subspecies available.

Patterns of genetic differentiation

Genomic differentiation (F_{ST}) was strongest between *H. sapho* vs. *H. hewitsoni* followed by *H. congener* vs. *H. eleuchia,* and Andean vs. Amazonian *H. antiochus* (average F_{ST} = 0.33, 0.26, and 0.07 respectively; Figure 2.4). *H. sara* vs. *H. leucadia* were the least differentiated pair (average F_{ST} = 0.05, Figure 2.4). In contrast, absolute divergence (D_{XY}) was highest in *H. sara* vs. *H. leucadia* while all other species pairs exhibited a similar degree of divergence (Figure 2.4). We also observed elevated F_{ST} values on the Z chromosome compared to autosomes in all but one comparison (*H. eleuchia* vs. *H. congener*; Figure 2.4). Interestingly, Chr4 shows lower F_{ST} values and increased D_{XY} values (in all comparisons but *H. sara* vs. *H. leucadia*). Only in *H. sapho* vs. *H. hewitsoni* Chr9 showed lower F_{ST} but did not show higher D_{XY} (Figure 2.4). Chr14 behaves like Chr9 but only in *H. eleuchia* and *H. congener* (Figure 2.4). The same pattern was true when we compared F_{ST} and D_{XY} between subspecies (Figures S2.30 – S2.33). Interestingly, F_{ST} values peaked at the end of Chr14 when comparing between subspecies of *H. congener*, in the same region that groups individuals by species and not sex in the local PCA (Figure S2.30).

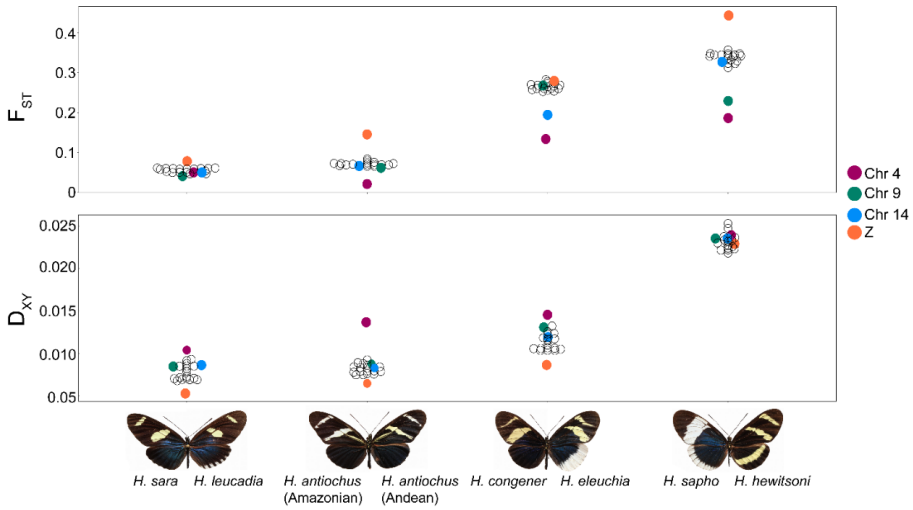

Figure 2.4 Genome-wide divergence (FST and DXY) between pairs of species in the *sara/sapho* clade. Each dot represents a chromosome, and chromosomes with evidence of

SA fusions are colour coded. Analyses for pairs of subspecies are shown in Figures S2.30 and S2.33.

Sex-specific differences in heterozygosity and coverage on multiple chromosome

The grouping by sex we observed in the local PCA and the trees exclusively of Chr4, Chr9 and Chr14 suggest possible fusions between these autosomes and the Z or W chromosome (females of *Heliconius* are ZW and males are ZZ). Because females of Lepidoptera lack crossing over, and their meiosis is achiasmatic, they do not recombine (55). This means that, if the Z chromosome is involved in the fusion (Figure 2.5A), only females would have unfused autosomes that would tend to accumulate mutations and/or structural variants leading to divergence from its (now fused) homologous. Alternatively, if the W chromosome is involved in the fusion (Figure 2.5B), the SA fusion would be restricted to females and the fused SA chromosome would tend to accumulate mutations and/or structural variants leading to divergence from its unfused homologous. Both scenarios would lead to high heterozygosity and low depth in females due to poor mapping (Figure 2.5). Consistently, Chr4, Chr9 and Chr14 showed striking sex-specific differences in heterozygosity in multiple species, supporting three sequential fusions of these chromosomes with the Z or W chromosomes (Figures 2.6A and 2.6B). The strongest difference in heterozygosity was observed on Chr4, where females of *H. eleuchia, H. congener, H. sapho, H. hewitsoni and H. antiochus* showed much higher heterozygosity than males (Figure 2.6B). In *H. eleuchia* and *H. congener*, females also showed very high heterozygosity on Chr14, and in *H. hewitsoni* and *H. sapho,* on Chr9 (Figure 2.6B). Differences between males and females were significant on these 3 chromosomes for all species of the sapho clade (Wilcoxon test, $p < 0.01$, Figures S2.34-2.38), except for *H. congener* and *H. hewitsoni* where differences could not be tested due to low sample size. Heterozygosity of Chr4 for females was also significantly higher than for the other chromosomes in *H. eleuchia*, *H. sapho*, *H. antiochus* (Wilcoxon test, $p < 0.01$, Figure S2.36A). The same was true for Chr14 in *H. eleuchia* and Chr9 in *H. sapho* (Wilcoxon test, $p < 0.01$, Figure S2.36A).

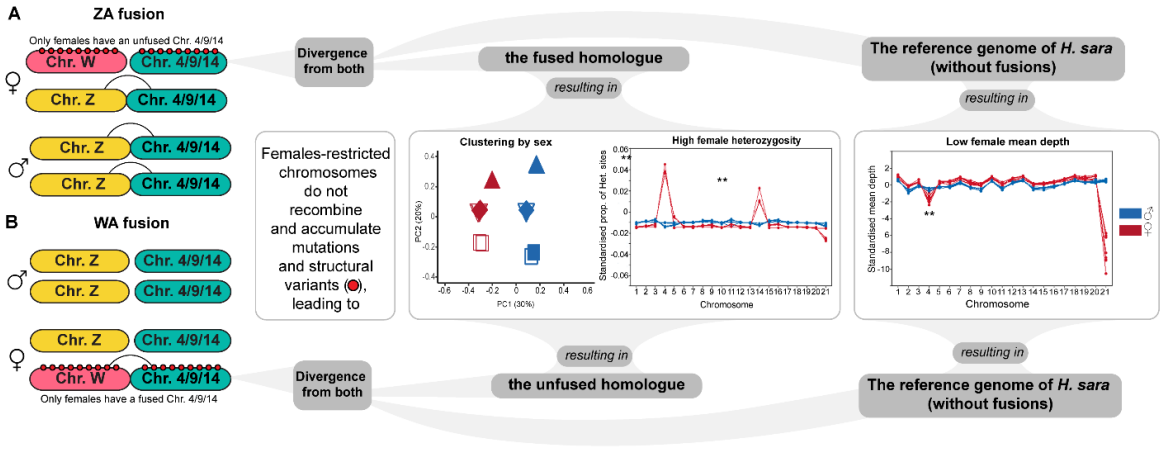

Figure 2.5 Scenarios of SA fusions involving either the W or Z chromosome. Scenario of A) ZA fusion and B) WA fusion for Chr4, Chr9 and Chr14 differentiating the expected pattern by sex. Yellow: Z chromosome. Pink: W chromosome. Green: autosome. Curved lines: fusions. Red dots: mutations and/or structural variants.

Females of *H. eleuchia*, *H. sapho*, *H. hewitsoni* and *H. antiochus* have reduced mean depth on chromosome 4, whereas the mean depth on that chromosome in males is normal (Figure 2.6B). These differences were significant in *H. eleuchia*, *H. sapho, H. antiochus* (Wilcoxon test, $p \leq 0.01$; Figure S2.34). However, this pattern was not true for Chr14 in *H. eleuchia* and *H. congener,* nor Chr9 for *H. sapho* and *H. hewitsoni* (Figures 2.6B and S2.35). The mean depth of Chr4 was also lower than that of all other chromosomes in females of *H. eleuchia*, *H. sapho* and *H. antiochus* (Wilcoxon test, $p \leq 0.01$; Figure S2.36A). However, this was not true for Chr14 and Chr9 (Figure S2.36A).

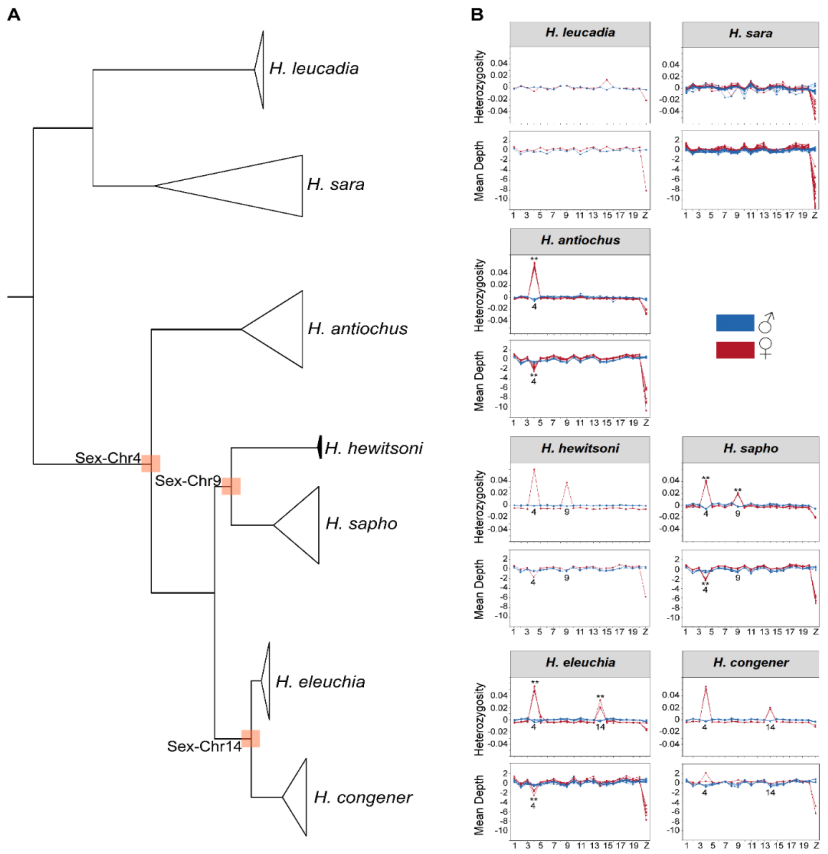

Figure 2.6 Genome-wide topology and patterns of heterozygosity and depth across the genome. A) Genome-wide topology with orange squares highlight clades where SA fusions were detected, and the SA fusion is specified. B) Standardised proportion of heterozygosity sites and standardised mean depth per chromosome in each species (**p<0.01). Each line corresponds to one individual.

The sliding window analyses on Chr4, Chr9 and Chr14 revealed that the excess heterozygosity in females is present in most windows along the entire chromosomes (i.e., it is not concentrated in a single continuous region; Wilcoxon test, p <0.01, Figures 2.7 and S2.37-S2.39). Also, mean depth values were significantly lower for females than males in most windows on Chr4 (*H. eleuchia, H. congener, H. sapho* and *H. antiochus*), Chr9 (*H. sapho*) and Chr14 (*H. eleuchia* and *H. congener*) (Wilcoxon test, p <0.01, Figures 2.7 and S2.37-S2.39). Interestingly, one female of *H. congener* has a region of eight windows on Chr4 with very high sequencing depth (Figure 2.7). There are also few windows on Chr4, Chr9 and Chr14 where females exhibited both lower heterozygosity and mean depth than males (Figure 2.7). *H. sara* and *H. leucadia* were the only species in the clade that did not show sex-specific

patterns in heterozygosity and mean depth in Chr4, Chr9 and Chr14 (Figures 2.6B and 2.7). Interestingly, the heterozygosity in females of *H. congener* dropped at the end of Chr14 to values similar to those of the males (Figure S2.39).

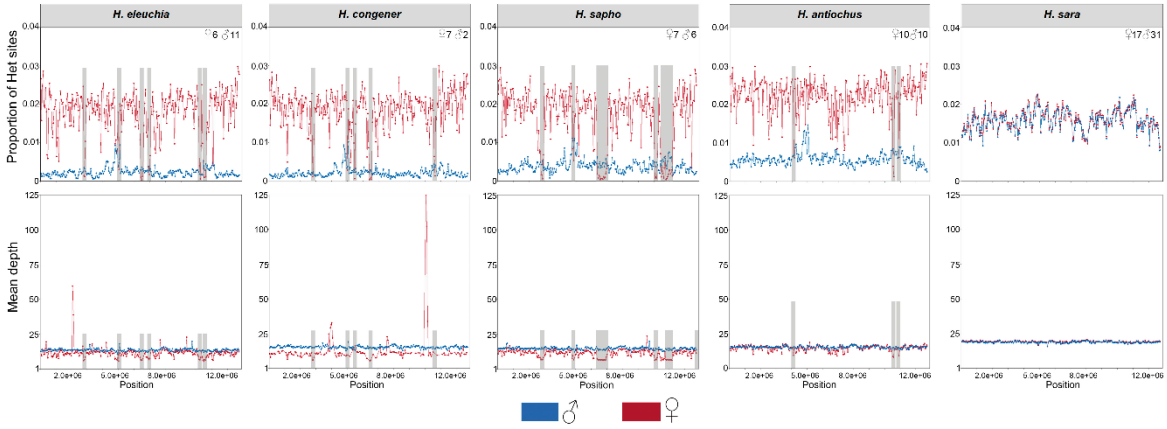

Figure 2.7 Patterns of heterozygosity and depth across chromosome 4. Proportion of heterozygosity sites and mean depth in 50Kb sliding windows in each species. Males are shown in blue and females in red, and their *n* is shown in the top right corner of each panel. Shadowed regions are those where females show very low sequencing depth and no heterozygous sites, suggesting that the female-limited haplotype diverged enough to become unmappable to the *H. sara* reference genome.

Discussion

We find evidence for the fusion of three autosomes to sex chromosomes in the evolution of the *sara/sapho* clade. The patterns of (i) sex-specific clustering in the phylogenies and local PCA, (ii) low F_{ST} on Chr4, Chr9 and Chr14, and (iii) high female heterozygosity seen on these chromosomes suggest fusions of a sex chromosome with these autosomes. As females are the heterogametic sex in butterflies and show no recombination, SA fusions should generate female-specific haplotypes that do not recombine with the unfused chromosomes. This results in high heterozygosity in females and increases variation within populations and reduces variation between them, as females share haplotypes with females from other populations, thus generating the observed reduced F_{ST} pattern. Chr4 shows sex-specific clustering and high female heterozygosity throughout the chromosome

in all five species of the *sapho* group, indicating that either the Z or W fused with Chr4 in the ancestor of the *sapho* group. In line with an old fusion and increasing degeneration, females show a very high proportion of heterozygous sites in most parts of the chromosome, whereas in other parts of this chromosome, females show very low sequencing depth and no heterozygous sites, indicating that the female-limited haplotype diverged enough to become difficult to map to the *H. sara* reference genome. Similar, but less strong patterns of female-biased heterozygosity and sex-specific clustering in phylogenies are found on Chr9 in *H. sapho* and *H. hewitsoni* and on Chr14 in *H. eleuchia* and *H. congener.* In line with younger fusions with only weak degeneration, the sequencing depth of females matches that of the males on those chromosomes, indicating that the female-specific haplotypes still map well to the *H. sara* genome. In summary, we find three independent SA fusions in the *sapho* clade, each shared by two to five species (Figure 2.1). It is unknown if the autosomes are fused with the Z or W chromosome, as both types of SA fusions would lead to female-limited haplotypes that no longer recombine (Chr4/9/14 in the ZA fusion and Chr W + Chr.4/9/14 in the WA fusion; Figure 2.5). Long-read sequencing data will be required to distinguish between these possibilities. No SA fusions have been reported previously for any *Heliconius* butterflies.

The finding of multiple SA fusions in the *sapho* group is particularly striking, since this group is known for its high number of chromosomes compared to all other *Heliconius* species. While most *Heliconius* species have 21 chromosomes, *H. eleuchia eleusinus* have 56-57 chromosomes, and *H. eleuchia* and *H. congener* have 37 and 33 chromosomes, respectively (54). This clearly suggests many chromosomal fission events, but surprisingly, our results show that SA fusions also occurred in this group. In all species, the sex-linked chromosomes show even patterns of increased heterozygosity and clustering by sex across, except for the end of Chr14 in *H. congener*. This indicates that the many chromosome fission events in the *sapho* group appear to have had little impact on the chromosomes that were fused with the sex chromosomes.

While we do not have evidence for an adaptive role of the SA fusions in the *sapho* group, the fact that SA fusions occurred three times and remained fixed in

multiple species, suggests that they are at least not deleterious, or that any deleterious effect is masked. In the latter scenario, low recombination around the SA fusions would force their sex-specific transmission and result in a permanent heterozygosity that protects against the expression of deleterious recessive mutations load and favours the accumulation of adaptive mutations (4). Alternatively, selection for suppressed recombination as a result of SA fusions could lead to sex-determining genes being tightly linked to sexually antagonistic loci on Chr4/9/14, thus coupling sex-specific beneficial alleles (1,7,9). Also, the SA fusions may have contributed to the particularly high diversification rate in this clade if they linked together barrier loci in regions with reduced recombination (12).

This is the first genomic study focused on the s*ara/sapho* clade. The inclusion of multiple species and subspecies of this clade from a broad geographic range allowed us to redefine some of the relations previously reported (34), and to identify the effect of geography in shaping diversity. The phylogenetic position we found for *H. antiochus* and *H. hewitsoni* contrasts with previous amplicon based phylogenies (32,34) but agrees with a recent whole genome phylogeny based on *de novo* genome assemblies (56), suggesting that the phylogenetic relations we describe for these two species are the most plausible. We also identified *cis* and *trans*-Andean lineages for *H. sara* and *H. antiochus*, as well as *H. congener* and *H. eleuchia* structured by the Andes. In addition, *H. sara* was the only species in which we identified Andean and Amazonian lineages. The driving forces behind the SA fusions identified here, as well as their role (if any) in the speciation or adaptation of this clade, remain unknown. Our study adds to a growing number of examples in butterflies that, in the future, may help to unravel the role of these rearrangements in the evolution of Lepidoptera and other organisms.

References

1. Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL. Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). PLOS Genet. 2009 Feb 20;5(2):e1000391.

2. Yoshido A, Sahara K, Marec F, Matsuda Y. Step-by-step evolution of neo-sex chromosomes in geographical populations of wild silkmoths, Samia cynthia ssp. Heredity. 2011 Apr;106(4):614–24.

3. Šíchová J, Ohno M, Dincă V, Watanabe M, Sahara K, Marec F. Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, *Leptidea amurensis*. Biol J Linn Soc. 2016 Jul;118(3):457–71.

4. Jay P, Tezenas E, Véber A, Giraud T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. PLOS Biol. 2022 Jul 19;20(7):e3001698.

5. Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, et al. A role for a neo-sex chromosome in stickleback speciation. Nature. 2009 Oct;461(7267):1079–83.

6. Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, et al. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. PLOS Genet. 2015 May 20;11(5):e1005237.

7. Sigeman H, Ponnikas S, Chauhan P, Dierickx E, Brooke M de L, Hansson B. Repeated sex chromosome evolution in vertebrates supported by expanded avian sex chromosomes. Proc R Soc B Biol Sci. 2019 Dec 4;286(1916):20192051.

8. Charlesworth D, Charlesworth B. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet Res. 1980 Apr;35(2):205–14.

9. Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. Evidence of a neo-sex chromosome in birds. Heredity. 2012 Mar;108(3):264–72.

10. Krátká M, Šmerda J, Lojdová K, Bureš P, Zedek F. Holocentric Chromosomes Probably Do Not Prevent Centromere Drive in Cyperaceae. Front Plant Sci [Internet]. 2021 [cited 2023 Mar 21];12. Available from: https://www.frontiersin.org/articles/10.3389/fpls.2021.642661

11. Pardo-Manuel de Villena F, Sapienza C. Female meiosis drives karyotypic evolution in mammals. Genetics. 2001 Nov;159(3):1179–89.

12. Guerrero RF, Kirkpatrick M. Local adaptation and the evolution of chromosome fusions. Evolution. 2014 Oct;68(10):2747–56.
13. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci U S A. 2013 May 7;110(19):E1743-1751.
14. Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, et al. Neo-sex chromosomes and adaptive potential in tortricid pests. Proc Natl Acad Sci U S A. 2013 Apr 23;110(17):6931–6.
15. Matsumoto T, Kitano J. The intricate relationship between sexually antagonistic selection and the evolution of sex chromosome fusions. J Theor Biol. 2016 Sep 7;404:97–108.
16. Schmid M, Feichtinger W, Steinlein C, Visbal García R, Fernández Badillo A. Chromosome banding in Amphibia. XXVIII. Homomorphic XY sex chromosomes and a derived Y-autosome translocation in Eleutherodactylus riveroi (Anura, Leptodactylidae). Cytogenet Genome Res. 2003;101(1):62–73.
17. Král J, Kořínková T, Krkavcová L, Musilová J, Forman M, Herrera IMÁ, et al. Evolution of karyotype, sex chromosomes, and meiosis in mygalomorph spiders (Araneae: Mygalomorphae). Biol J Linn Soc. 2013 Jun 1;109(2):377–408.
18. Flores SV, Evans AL, McAllister BF. Independent Origins of New Sex-Linked Chromosomes in the melanica and robusta Species Groups of Drosophila. BMC Evol Biol. 2008 Jan 29;8:33.
19. Bressa MJ, Papeschi AG, Vítková M, Kubícková S, Fuková I, Pigozzi MI, et al. Sex chromosome evolution in cotton stainers of the genus Dysdercus (Heteroptera: Pyrrhocoridae). Cytogenet Genome Res. 2009;125(4):292–305.
20. Mongue AJ, Nguyen P, Voleníková A, Walters JR. Neo-sex Chromosomes in the Monarch Butterfly, *Danaus plexippus*. G3 GenesGenomesGenetics. 2017 Oct 1;7(10):3281–94.
21. Smith DAS, Gordon IJ, Traut W, Herren J, Collins S, Martins DJ, et al. A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. Proc R Soc B Biol Sci. 2016 Jul 27;283(1835):20160821.

22. Šíchová J, Voleníková A, Dincă V, Nguyen P, Vila R, Sahara K, et al. Dynamic karyotype evolution and unique sex determination systems in Leptidea wood white butterflies. BMC Evol Biol. 2015 Dec;15(1):89.
23. Carabajal Paladino LZ, Provazníková I, Berger M, Bass C, Aratchige NS, López SN, et al. Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea. Barluenga M, editor. Genome Biol Evol. 2019 Apr 1;11(4):1307–19.
24. Picq S, Lumley L, Šíchová J, Laroche J, Pouliot E, Brunet BMT, et al. Insights into the Structure of the Spruce Budworm (*Choristoneura fumiferana*) Genome, as Revealed by Molecular Cytogenetic Analyses and a High-Density Linkage Map. G3 GenesGenomesGenetics. 2018 Aug 1;8(8):2539–49.
25. Yoshido A, Marec F, Sahara K. Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG) n telomeric probe in some species of Lepidoptera. Chromosoma. 2005 Aug;114(3):193–202.
26. Lucek K, Augustijnen H, Escudero M. A holocentric twist to chromosomal speciation? Trends Ecol Evol. 2022 Apr;S0169534722000854.
27. White MJD. Animal Cytology and Evolution [Internet]. 3rd ed. 1977 [cited 2022 Dec 6]. Available from: https://www.cambridge.org/co/academic/subjects/life-sciences/cell-biology-and-developmental-biology/animal-cytology-and-evolution-3rd-edition, https://www.cambridge.org/co/academic/subjects/life-sciences/cell-biology-and-developmental-biology
28. Hill J, Rastas P, Hornett E, Neethiraj R, Clark N, Morehouse N, et al. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci Adv. 2019 Jun 12;5:eaau3648.
29. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. Major Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. G3 GenesGenomesGenetics. 2016 Mar 1;6(3):695–708.
30. Brown KS. The biology of *Heliconius* and related genera. Annu Rev Entomol. 1981;26(4):27–56.

31. Jiggins CD, Lamas G. The Ecology and Evolution of *Heliconius* Butterflies [Internet]. Oxford University Press; 2016 [cited 2022 Jul 18]. Available from: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199566570.001.0001/acprof-9780199566570

32. Beltrán M, Jiggins CD, Brower AVZ, Bermingham E, Mallet J. Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. Biol J Linn Soc. 2007 Oct 1;92(2):221–39.

33. Engler-Chaouat HS, Gilbert LE. De novo synthesis vs. sequestration: negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. J Chem Ecol. 2007 Jan;33(1):25–42.

34. Kozak KM, Wahlberg N, Neild AFE, Dasmahapatra KK, Mallet J, Jiggins CD. Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic *Heliconius* Butterflies. Syst Biol. 2015 May 1;64(3):505–24.

35. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020 Feb;17(2):155–8.

36. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv; 2012 [cited 2023 Mar 6]. Available from: http://arxiv.org/abs/1207.3907

37. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021 Feb 16;10(2):giab008.

38. Chow W, Brugger K, Caccamo M, Sealy I, Torrance J, Howe K. gEVAL — a web-based browser for evaluating genome assemblies. Bioinformatics. 2016 Aug 15;32(16):2508–10.

39. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, et al. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLOS Comput Biol. 2019 Aug 21;15(8):e1007273.

40. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021 Oct 1;38(10):4647–54.

41. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. ArXiv; 2013 [cited 2022 Jul 29]. Available from: http://arxiv.org/abs/1303.3997

42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

43. Broad I. Picard Toolkit [Internet]. Broad Institute; 2019 [cited 2022 Jul 29]. Available from: https://github.com/broadinstitute/picard

44. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491–8.

45. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

46. Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. Mol Biol Evol. 2018 Jun 1;35(6):1489–506.

47. Stamatakis A, Hoover P, Rougemont J. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Renner S, editor. Syst Biol. 2008 Oct 1;57(5):758–71.

48. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Dec 1;4(1):s13742-015-0047–8.

49. Purcell S, Chang CC. PLINK: Whole genome data analysis toolset [Internet]. 2007 [cited 2022 Jul 29]. Available from: www.cog-genomics.org/plink/2.0/

50. Li H, Ralph P. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. Genetics. 2019 Jan 1;211(1):289–304.

51. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Viena, Austria; 2013. Available from: http://www.R-project.org/

52. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009 Sep;19(9):1655–64.

53. Rueda-M N, Salgado-Roa FC, Gantiva-Q CH, Pardo-Díaz C, Salazar C. Environmental Drivers of Diversification and Hybridization in Neotropical Butterflies. Front Ecol Evol [Internet]. 2021 [cited 2022 Oct 31];9. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2021.750703

54. Brown KS, Emmel TC, Eliazar PJ, Suomalainen E. Evolutionary patterns in chromosome numbers in neotropical Lepidoptera. I. Chromosomes of the Heliconiini (family Nymphalidae: subfamily Nymphalinae). Hereditas. 1992;117(2):109–25.

55. Suomalainen E, Cook LM, Turner JRG. Achiasmatic oogenesis in the Heliconiine butterflies. Hereditas. 1973;74(2):302–4.

56. Cicconardi F, Milanetti E, Castro ÉCP de, Mazo-Vargas A, Belleghem SMV, Ruggieri AA, et al. Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies [Internet]. bioRxiv; 2022 [cited 2022 Oct 29]. p. 2022.08.12.503723. Available from: https://www.biorxiv.org/content/10.1101/2022.08.12.503723v1

Chapter 3

Chemical defence variation in *Heliconius* butterflies: testing the role of mimicry rings and ecoregions

Introduction

Cyanogenic glucosides (CNglcs) are secondary metabolites found in many plants where they usually play a role of protection from herbivores (1). The presence of CNglcs in animals seems to be restricted to arthropods (2), particularly to Coleoptera, Heteroptera and Lepidoptera (3,4) which obtain them via *de novo* synthesis, or more commonly, via sequestration from host plants (5,6). Insects usually use CNglcs as deterrents due to their bitterness and as defence due to the release of hydrogen cyanide (HCN) (4,7). These defences are usually coupled with bright colour patterns (warning signals) that local predators learn to avoid, which is known as aposematism (8). Because benefits of aposematism to preys increase with population size, multiple prey species within an area may evolve towards a single local phenotype thereby leading to local Müllerian mimicry rings (9). The resulting convergence in colour patterns allows co-mimics to share the cost of teaching predators and lower the *per capita* predation risk (10,11). Despite the theoretical premise that convergence should favour the establishment of a single local Müllerian ring, there is an immense mimetic diversity in nature even at the population level, which leads to local polymorphisms. However, the latter is possible if toxic compounds are deterrent enough to predators (12).

Variation in chemical defences within and between prey populations both in terms of chemical composition and concentration has been widely documented (13–16). Although the drivers of this variation are still unknown, frequency dependent selection and genetic variability may broadly contribute to the toxicity spectrum (17,18). In species that only sequester CNglcs, toxicity may be constrained by geographic and temporal availability of host plants (19,20), variation in the

concentration and/or availability of CNglcs in the host plant (21,22), and detoxification capacity of the prey to overcome the chemical arsenal of the host plant (23). In contrast, in prey species that *de novo* synthesise, variation in the cyanogenic profile may be affected by age, reproductive state, resource optimisation, as well as availability of precursors for CNglcs biosynthesis (24,25). Also, because the protection that CNglcs provide to a prey does not increase with toxin concentration (12,26), variation in toxicity may be the result of simple drift due to cyanogenic profiles being selectively neutral (13,27). In this case, the effectiveness of the warning signal and/or the mimicry dynamics would rely upon the frequency of the signal rather than the level of toxicity (10). Therefore, it is important to determine whether variation in toxicity is ubiquitous in communities of mimetic species that are chemically defended, and if so, investigate the causes behind such variation and its effects on the mimicry dynamics.

Mimetic butterflies of the genus *Heliconius* are chemically defended by cyanogens that they obtain either via *de novo* synthesis or sequestration from their host plants (28). Biosynthesis is the most used strategy by these butterflies (29), and it seems to be more important in generalist (28). In contrast, monophagous specialists mostly sequester (28,30). For example, some species in the *sara/sapho* clade exclusively feed on plants of the subgenus *Astrophea* from each they obtain CNglcs, and consequently, their cyanogenic profile contains virtually no biosynthesised CNglcs (28). Multiple studies have addressed cyanogenic variation in *Heliconius.* One compares sympatric but non-mimetic species, co-mimics, and mimicry rings in a single locality in Peru (16). Another compares between colour patterns and subspecies in different countries (31). A third one focuses on the species *H. erato* and compares between geographic populations and environmental gradients (6). However, to date, no study has tested the generality of the patterns described in those studies, especially in mimetic communities that occur in different Neotropical ecoregions (32).

In this study we investigated whether there is variation in the cyanogenic profile and concentration in *Heliconius* from multiple localities across Colombia that encompass six out of the seven Neotropical ecoregions. We compared the CNglc of

Heliconius butterflies: (i) between phylogenetic clades, (ii) between and within species, (iii) between sympatric but non mimetic species, and (iv) between and within mimicry rings. We also investigated whether cyanogenic variation was explained by phylogenetic relationships, host plant specialization, ability to biosynthesize or sequester CNglcs, or the mimicry ring to which a species belongs.

Materials and methods

Sample collection and metabolites extraction

We collected 240 individuals of *Heliconius* from 20 species across Colombia (Table S3.1 and Figure 3.1). We immediately preserved the right wing, half of the thorax, and the abdomen of each individual in a tube containing ~1 mL of 100% methanol to preserve cyanogenic glucosides (CNglcs). The remaining tissue was preserved in DMSO, and left wings were stored in glassine envelopes. Long term storage was done keeping the tubes at -80°C until further processing. All individuals were deposited in the 'Colección de Artrópodos de la Universidad del Rosario – CAUR 229'.

Samples in methanol were evaporated at room temperature using the Savant Automatic Environmental SpeedVac System AES1010 until the tissue was fully dried. Then, the tissue was homogenised using a porcelain mortar and a pestle, the resulting powder was added into a previously weighed Eppendorf tube filled with 800 µl of 80% methanol, and the mix was vigorously vortexed. Each extract was centrifuged at 14,000 *g* for 5 min and the supernatant was collected in a glass vial. We then filtered 45 µl of the supernatant and made a 50X (v/v) dilution in ultra-pure water. Individual extracts were analysed by Liquid Chromatography - Mass Spectrometry (LC-MS) that was conducted either in a LC-Orbitrap (Bruker Daltonics, Bremen, Germany) or a LC-qToF (Bruker Daltonics, Bremen, Germany) mass spectrometer.

Liquid Chromatography-Mass Spectrometry (LC-MS/MS) and chemical analyses

LC-MS was conducted as previously described for both the LC-Orbitrap-MS (28) and for the LC-qToF-MS (33). In both cases, LC-MS data was analysed using the Bruker Compass DataAnalysis 4.3 software (Bruker Daltonics, Bremen, Germany). CNglcs were identified as formic adducts [M + CH_2O_2] in samples analysed by LC-qToF-MS, whereas in those analysed by LC-Orbitrap-MS, CNglcs were identified as sodium adducts [M + Na^+] (Table S3.2). In all cases, the concentration of each CNglc was calculated using the peak area ratio of the analyte to the internal standard (amygdalin) using a regression equation generated from a five-point calibration curve for the standard (concentrations of 0.5, 1.0, 2.0. 5.0 and 20.0 ng/mL). The concentration of each CNglc is reported in µg of CNglc/mg of dry butterfly weight. To make sure our data was comparable despite it being generated in two different MS, we tested whether the amount of CNglcs resulting from the LC-qToF-MS were different from those from the LC-Orbitrap-MS with a Mann-Whitney U-test (as data was not normal and not suitable for transformation).

We organised the results per individual into two sets: (i) concentration of all CNglcs in the sample - referred to as '*CNglcs total*', and (ii) concentration of each of the CNglcs identified in the sample - referred to as '*CNglc profile*'.

Statistical analyses

We first tested for differences in mean '*CNglcs total*' between species with a non-parametric Kruskal-Wallis followed by a post-hoc Dunn test. We adjusted *p*-values for multiple comparisons using the method *false* discovery rate (*FDR*). We then compared between sympatric species and between mimetic pairs from the same location using a Kruskal-Wallis. Additionally, we tested for intraspecific differences within each species. We conducted our analyses and created figures using the ggstatsplot package in R software (34).

We later performed a multivariate analysis of variance (PERMANOVA) in R to test for differences in '*CNglc profile*' species, sympatric species and mimetic pairs from the same location, and for intraspecific differences within each species. The

latter were set as predictor variables whilst '*CNglc profile*' data was transformed into a Euclidean dissimilarity matrix and used as the dependent variable. The analysis was carried out with the *adonis2* function in the R package Vegan (35), with 1000 permutations. The post-hoc analysis was conducted using the *pairwise.adonis* function and the *FDR* method also in Vegan (35). As before, we compared between sympatric species and between mimetic pairs from the same location, as well as within species, using PERMANOVA. Heatmaps were plotted using the function *heatmap.2* in the R package gplots (36).

To investigate whether the '*CNglc profile'* of *Heliconius* is explained by their phylogeny, we assessed the phylogenetic signal with a phylogenetically controlled MANOVA (PhyloMANOVA) using the *aov.phylo* function from the package *geiger* (37) with 10,0000 randomizations. Then, we examined the variation of CNglcs in a non-metric multidimensional scaling (NMDS) calculating a Bray-Curtis dissimilarity index between individuals using *metaMDS* function in the *vegan* (35). We calculated NMDS for dimensions ranging from 1 to 10 in order to find the best number of dimensions, and used the *NMDS.scree()* function to check the associated stress values as a measure of goodness-of-fit (https://ourcodingclub.github.io/tutorials/ordination/#section6). We next used the *envfit* function of *vegan* to overlap pupal and non-pupal mating clades and phylogenetic clades onto the NMDS plots without disrupting the original ordinations (999 random permutations). This generates R^2 and significance values that reflect whether these clades are associated with the '*CNglc profile'*. We tested clade differences in the two main NMDS axes with a non-parametric Kruskal-Wallis followed by a post-hoc Dunn test. F*alse* discovery rate (*FDR*) correction for multiple testing was used to adjust *p*-values for multiple comparisons.

We also examined whether the '*CNglc profile'* of *Heliconius* is influenced by: (i) ability to biosynthesize or sequester CNglcs, (ii) host plant specialization, and (iii) mimicry ring. To test whether the ability to biosynthesize or sequester CNglcs is associated with the '*CNglc profile'* we first quantified biosynthesized compounds (linamarin, lotaustralin, and epilotaustralin) and sequestered compounds (gynocardin, dihydrogynocardin, tetraphyllin B, epivolkenin, and deidaclin) per

individual. We then categorized each individual as “synthesizer” when biosynthesized compounds were >50%, “catcher” when sequestered compounds were >50%, or “synthesizer + catcher” when biosynthesized and sequestered compounds were in similar proportion. To test whether host plant specialization is associated with ‘*CNglc profile’* we classified individuals based on their specialization in the following host plants: *Decaloba*, *Astrophea*, or *Passiflora*.

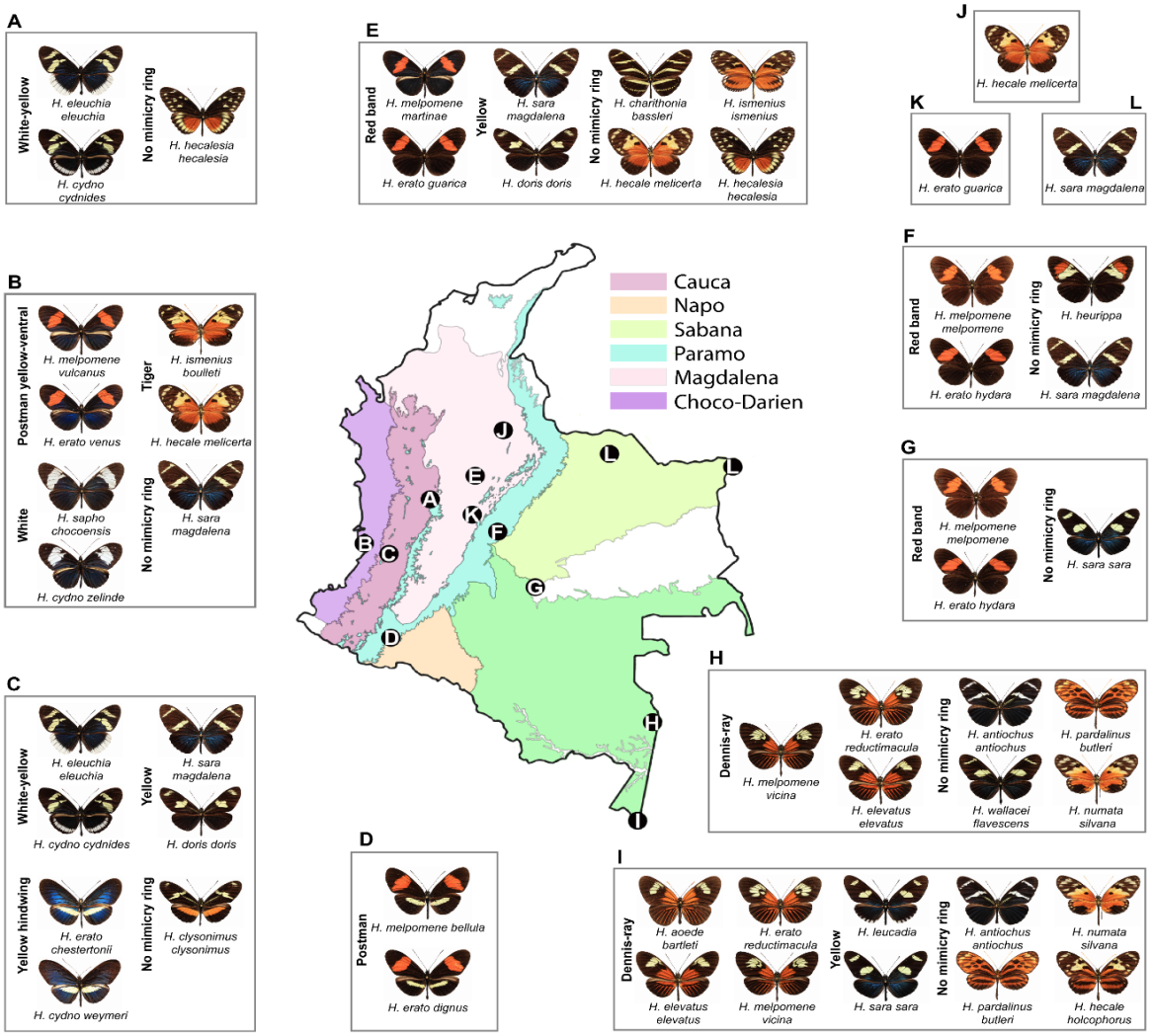

Figure 3.1 Geographic distribution of individuals included in this study. We collected butterflies in 12 locations across Colombia: (A) Caldas, (B) Buenaventura, (C) Cauca Valley, (D) Putumayo, (E) Boyacá, (F) Meta, (G) Guaviare, (H) Amazonas-Pedrera, (I) Amazonas-Puerto Nariño, (J) Santander, (K) Cundinamarca, and (L) Arauca and Vichada. The species collected in each location are shown, as well as the mimicry ring they participate in (unless otherwise indicated).

Results

Sample collection

We collected 240 individuals from 13 localities throughout Colombia, which led to include nine *Heliconius* mimicry rings: (i) dennis-ray, (ii) yellow, (iii) red band, (iv) postman, (v) tiger, (vi) white-yellow, (vii) white, (viii) postman yellow-ventral, and (ix) yellow hindwing (Figure 3.1 and Table S3.1). Although we sampled some *Heliconius* that participate in mimicry rings with members of other tribes or even moths, these non-*Heliconius* mimics were not included in our analyses (Figure 3.1). Also, in some localities we sampled a single species of *Heliconius* either because it does not participate in a mimicry ring or because we did not collect the mimic (Figure 3.1).

Quantification of all CNglcs combined ('CNglcs total')

Levels of CNglcs obtained by LC-qToF-MS and LC-Orbitrap-MS were not different (Mann-Whitney U-test, $p > 0.05$; Figure S3.1), and thus, we proceeded to combine the data. We found that only a few pairs species of *Heliconius* differ in their '*CNglcs total*' (out of 190 comparisons, 7 exhibited differences), with *H. sara* vs. *H. charithonia*, *H. sara* vs. *H. erato, H. sara* vs. *H. melpomene, H. sara* vs. *H. cydno, H. sara* vs. *H. hecale, H. sara* vs. *H. elevatus* and *H. antiochus* vs. *H. cydno* being the most different (Dunn post-hoc test; $p < 0.05$, Figure 3.2A). *H. eleuchia* was the species with the highest median concentration of CNglcs, followed by *H. sara* and *H. antiochus* (Figure 3.2A), meaning that the most toxic species in our sampling are in the *sara/sapho* clade.

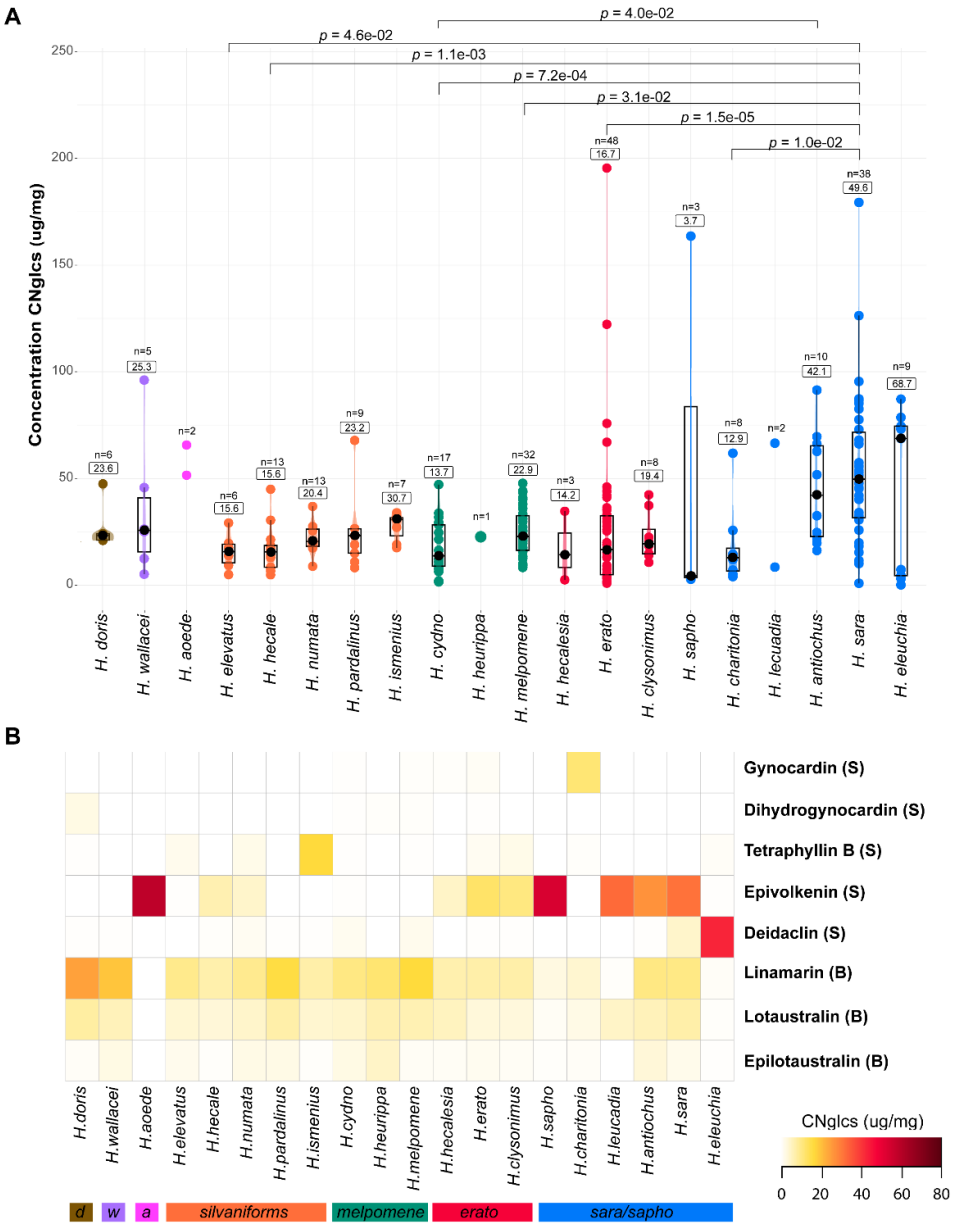

Figure 3.2 Quantification of CNglcs in *Heliconius*. (A) Quantification of all CNglcs combined or *'CNglcs total'*. Phylogenetic clades are colour coded, from left to right, as follows: *doris* (brown; 1 species), *wallacei* (purple; 1 species), *aoede* (pink; 1 species), *silvaniforms* (orange; 5 species), *melpomene* (green; 3 species), *erato* (red; 3 species), and *sara/sapho* (blue; 6 species). The number of individuals included per species (n) is indicated on top of each box, as well as the mean concentration of *CNglcs total* per species (numbers in squares), except when the number of individuals was less than 3. Significantly different comparisons ($\alpha < 0.05$) are shown. (B) Quantification of each CNglc or '*CNglc profile*' per species, where the mean concentration (ug/mg) of each compound is colour coded. The phylogenetic clade to which each species belongs is indicated at the bottom following the same colour code as in A. The *CNglcs* quantified are shown at the right, indicating whether they are biosynthesized (B) or sequestered (S).

Overall, we found sympatric species being similar in overall toxicity. First, sympatric but non-mimetic species were similar in toxicity except for *H. m. vulcanus* vs. *H. a. antiochus* (Buenaventura, Figure 3.3B top panel), *H. e. eleuchia* vs. *H. c. weymeri* (Cauca Valley, Figure 3.3C top panel), *H. e. hydara* vs. *H. s. magdalena* (Meta, Figure 3.3F top panel), and *H. e. reductimacula* vs. *H. a. antiochus* and *H. e. reductimacula* vs. *H. w. flavescens* (Pedrera-Amazonas, Figure 3.3H top panel). Second, most mimetic pairs do not differ in their overall toxicity except for *H. e. hydara* and *H. m. melpomene* in Guaviare (Figure 3.3G top panel). In contrast, we observed that the overall toxicity of two mimicry rings varied geographically (Figure S3.2). Specifically, the *white-yellow* mimicry ring was 17X more toxic in Cauca Valley than in Caldas (Figure S3.2A), while the *red-band* ring was more toxic in Cundinamarca than in any other locality (Figure S3.2B).

We also observed geographic variation in toxicity in *H. sara*, *H. erato*, and *H. melpomene* (Figure 3.4 and S3.3). In *H. sara,* for example, individuals from Buenaventura were the most toxic, while those from Puerto Nariño-Amazonas and Boyacá were the least (Figure 3.4A top panel). In *H. erato,* populations from Cundinamarca, Buenaventura and the Cauca Valley were significantly more toxic than populations from the jungle (Amazonas, Guaviare and Putumayo; Figure 3.4B top panel). In *H. melpomene,* individuals from Meta had the highest toxicity, which differed from those from Pedrera-Amazonas and Boyacá (Figure 3.4C top panel). In those species were we only sampled two populations, we observed significant differences in toxicity only for *H. eleuchia,* with individuals from Caldas being less toxic than those from the Cauca Valley (Figure 3.4D top panel and Figure S3.3). Interestingly, *H. cydno* did not exhibit significant geographic variation in toxicity despite this species participates in several mimicry rings (Figure 3.4E top panel).

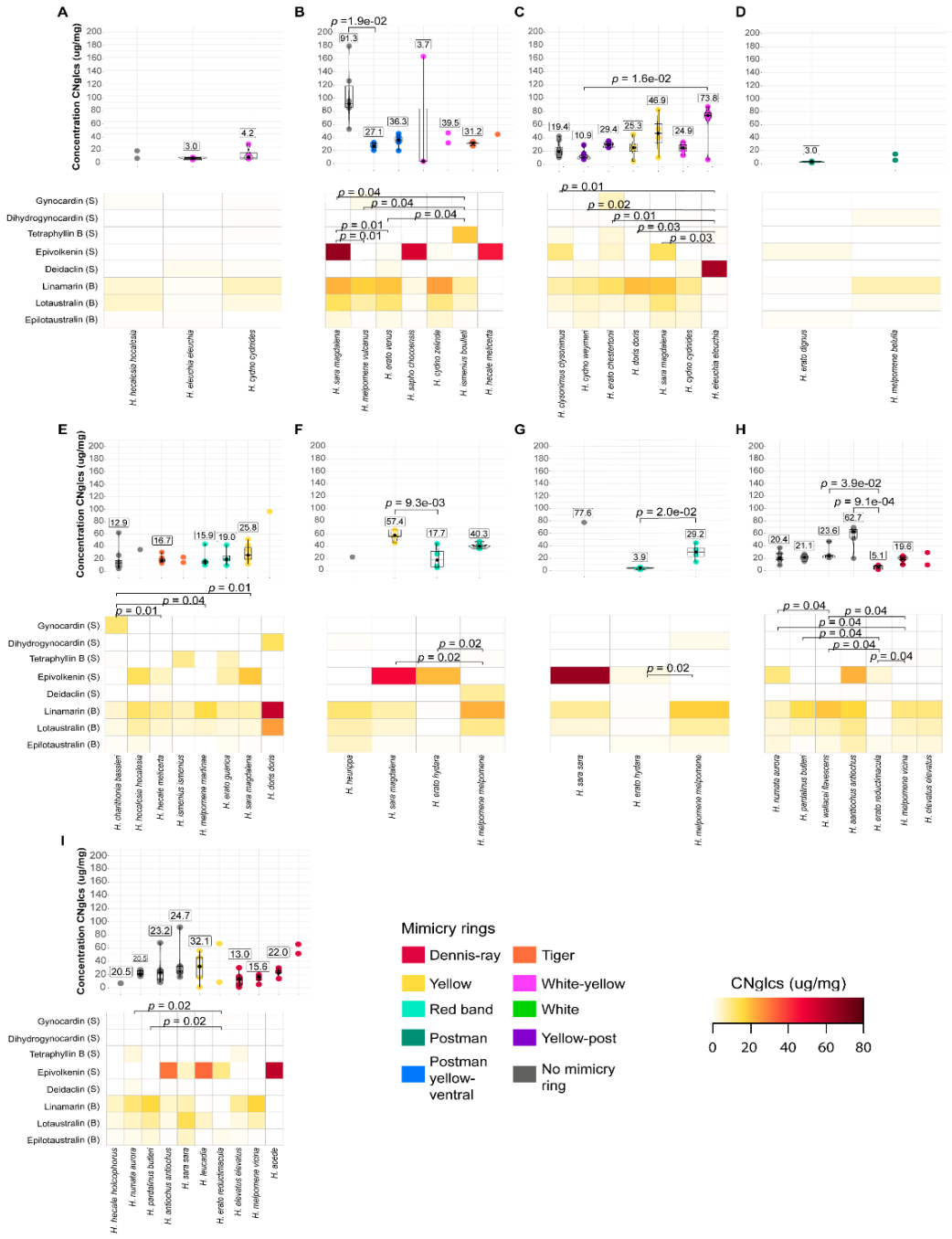

Figure 3.3 Quantification of cyanogenic glucosides per species and locality. (A) Caldas, (B) Buenaventura, (C) Cauca Valley, (D) Putumayo, (E) Boyacá, (F) Meta, (G) Guaviare, (H) Amazonas-Pedrera, (I) Amazonas-Puerto Nariño. Boxplots on the top of each panel show the mean concentration (numbers in squares) of all CNglcs combined or *'CNglcs total'* per species. The mimicry ring to which each species belongs is also indicated at the bottom using a colour code. Matrices on the bottom of each panel show the quantification of each CNglcs or '*CNglc profile*' per species, where the mean concentration (ug/mg) of each compound is colour coded from pale yellow to dark red. The *CNglcs* quantified are shown at the right,

indicating whether they are biosynthesized (B) or sequestered (S). Significantly different comparisons ($\alpha < 0.05$) are shown.

Quantification of each CNglc ('CNglc profile')

When we compared each of the nine CNglcs between the 20 species sampled (190 comparisons) we found 81 comparisons as significantly different (PERMANOVA $R^2 = 0.283$, p = 0.001, Table S3.3; pairwise adonis $p > 0.05$, Table S3.4). Interestingly, 66% of these significant comparisons included species of the *sara/sapho* clade, which characterised by having a higher concentrations of sequestered CNglcs (gynocardin, dihydrogynocardin, tetraphyllin B, epivolkenin, and deidaclin; Figure 3.2B) compared to biosynthesized CNglcs (linamarin, lotaustralin and epilotaustralin). Within this clade we found *H. sara*, *H. antiochus*, *H. leucadia* and *H. sapho* having larger amounts of epivolkenin, while *H. eleuchia* contained more deidaclin, and *H. charithonia* contained more gynocardin (Figure 3.2B). Additionally, although all members of the clade contained biosynthesized CNglcs, *H. eleuchia* and *H. sapho* showed only traces of these compounds (Figure 3.2B). In contrast, species of the *erato* clade exhibited a more uniform distribution of sequestered and biosynthesized CNglcs (Figure 3.2B), while species of the *melpomene, silvaniform* and *wallacei* clades contained mostly biosynthesized CNglcs (Figure 3.2B). Species from the *silvaniform* clade had higher levels of linamarin except for *H. ismenius*, which contained more Tetraphyllin B (Figure 3.2B). We unexpectedly observed *H. numata* and *H. hecale* having traces of the sequestered CNglc epivolkenin (Figure 3.2B). Finally, like most of the species in the *sara/sapho* clade, *H. aoede* had high concentrations of epivolkenin (Figure 3.2B).

Our results show that, in general, sympatric species are similar in terms of individual CNglc profiles (Figure 3.3 bottom panels and Table S3.3). However, there were some exceptions to this pattern. First, the only mimetic pair that significantly differed in CNglcs was *H. erato* and *H. melpomene* in Meta, Guaviare, and Amazonas (PERMANOVA $p < 0.05$, Figures 3.3F-3.3H, bottom panel). Second, when comparing between sympatric but non-mimetic species we found some CNglc being specific to a given species or in concentrations higher than in any other co-

occurring species (p < 0.05, Figure 3.3 bottom panels). For example, in Buenaventura tetraphyllin B was exclusive to *H. ismenius* while epivolkenin was exceptionally more abundant in *H. sara* (Figure 3.3B bottom panel). In the Cauca Valley, epivolkenin was highly abundant only in *H. eleuchia* (Figure 3.3C bottom panel). In Boyacá, gynocardin was exclusive to *H. charithonia* (Figure 3.3E bottom panel). In Meta, deidaclin was only found in *H. melpomene* while Linamarin was also highly abundant in this species (Figure 3.3F bottom panel). In Amazonas-Pedrera, the concentration of Linamarin was high only in *H. wallacei*, while *H. erato* contained negligible amounts of any CNglc (Figures 3.3H-3.3I bottom panel). We consistently observed across localities that the most toxic species were those with high concentrations of epivolkenin: (i) *H. sara* in Buenaventura, Meta, and Guaviare (Figures 3.3B, 3.3F and 3.3G bottom panel), (ii) *H. aoede*, *H. leucadia*, and *H. antiochus* in the Amazonas (Figure 3.3I bottom panel), and (iii) *H. hecale* and *H. sapho* in Buenaventura (Figure 3.3B bottom panel). There were only two exceptions to this general observation. The first, *H. doris* in Boyacá, that was highly toxic due to high concentrations of Linamarin (Figure 3.3E bottom panel), and *H. eleuchia* in the Cauca Valley, whose high toxicity was due to high concentrations of deidaclin (Figure 3.3G bottom panel).

CNglc profiles also varied geographically in *H. sara*, *H. erato* and *H. eleuchia* (p < 0.05, Table S3.3). First, epivolkenin was highly abundant in *H. sara* from Buenaventura, Guaviare, Arauca, and Meta whereas Deidaclin was more abundant in populations from Vichada and Arauca (Figure 3.4A bottom panel). Second, *H. erato* showed a remarkable geographic variation in its toxicity. Populations from Cundinamarca and Meta had more epivolkenin (sequestered), while populations from Buenaventura had more biosynthesized CNglcs. In contrast, populations from the Cauca Valley and Boyacá had uniform amounts of sequestered and biosynthesized CNglcs. Interestingly, populations from Amazonas, Guaviare and Putumayo were much less toxic than any other population (p < 0.05, Figure 3.4B bottom panel). Third, *H. eleuchia* from the Cauca Valley had high levels of deidaclin whereas populations from Caldas had low amounts of any CNglc (p > 0.05, Figure 3.4D bottom panel). Finally, no significant differences in individual CNglc profiles

were found between populations of *H. melpomene* or between populations of *H. cydno* ($p > 0.05$, Figures 3.4C and 3.4D bottom panels).

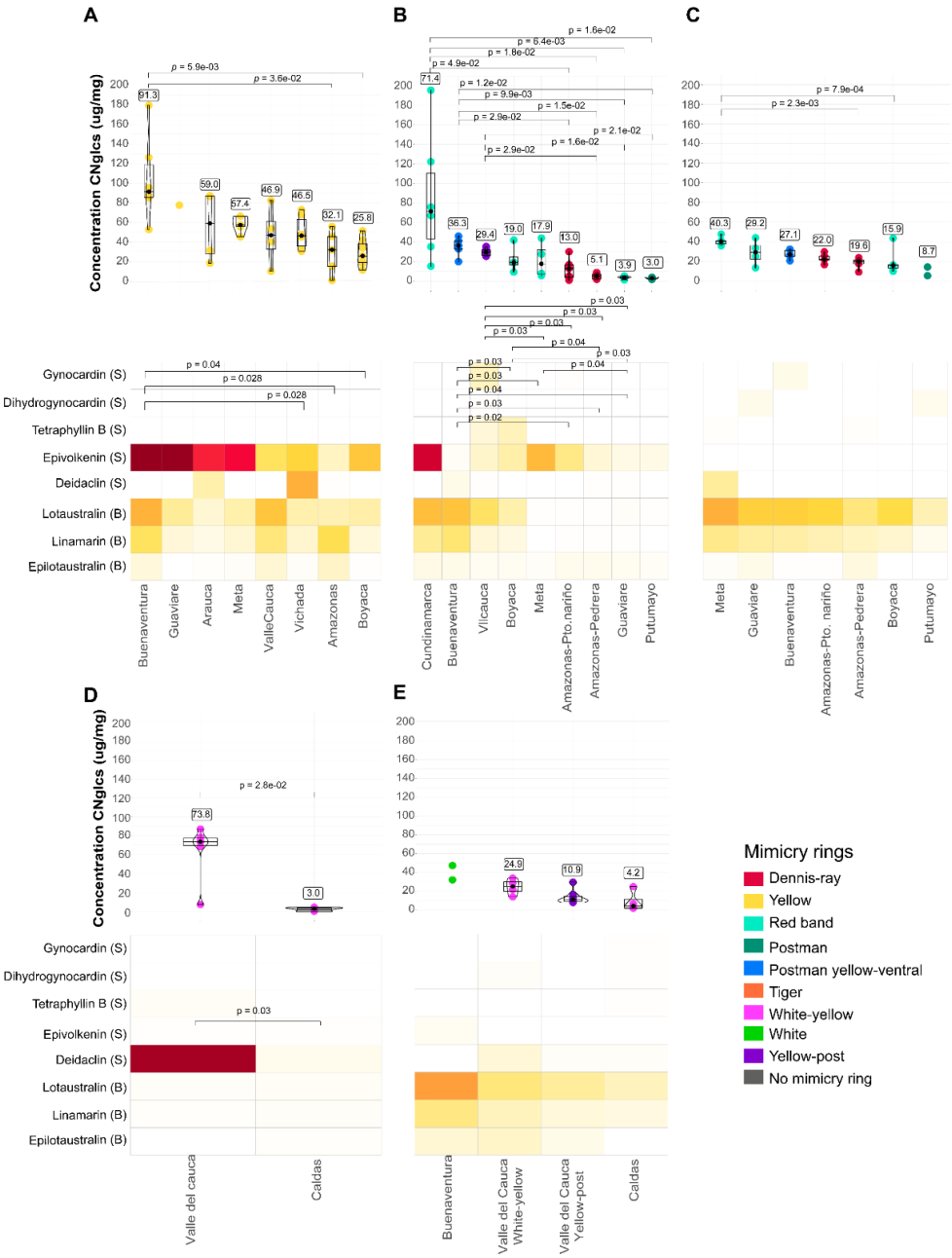

Figure 3.4 Intraspecific variation in the concentrations of cyanogenic glycosides. (A) *H. sara*, (B) *H. erato*, (C) *H. melpomene*, (D) *H. eleuchia*, (E) *H. cydno*. Boxplots on top of each panel show the mean concentration (numbers in squares) of all CNglcs combined or *'CNglcs total'* per locality (in µg/mg). The mimicry ring to which each species belongs is also indicated at the bottom using a colour code. Matrices on the bottom of each panel show the quantification of each CNglcs or '*CNglc profile*' per locality, where the mean concentration

(ug/mg) of each compound is colour coded from pale yellow to dark red. The *CNglcs* quantified are shown at the right, indicating whether they are biosynthesized (B) or sequestered (S). Significantly different comparisons ($\alpha < 0.05$) are shown.

The above differences did not hold when phylogeny was considered, suggesting that phylogenetic relationships influence the '*CNglcs profile'* (PhyloMANOVA F = 1573.4, df = 18, *p*= 0.000381, *p* given phy = 1). Consistently, we found that pupal, non-pupal, and phylogenetic clades are associated with the *CNglcs profile* (*p* < 0.05). However, most of the variation in *CNglcs* is due to differences between ability to biosynthesize or sequester cyanogenic glucosides of the individuals, followed by phylogenetic, pupal/non-pupal clades and host plant specialization (R^2 = 0.39, R^2 = 0.28, R^2 = 0.22, R^2 = 0.21 respectively; Figures 3.5 and 3.6A), rather than to differences among mimicry rings (R^2 = 0.14; Figures 3.6B and S3.4). The composition of CNglcs of the clades silvaniforms, melpomene, wallacei, and doris was different from that of the clades erato, *sara/sapho*, and aoede (p<0.05; Figure 3.5A). Only two individuals (one *H. hecale* and one *H. numata*) were different from other silvaniforms as they contained high amounts of epivolkenin (Figure 3.5A). We also found high variation among the pupal mating clades with several of these individuals falling into the non-pupal mating clades (p<0.05; Figure 3.5B). In contrast, only four individuals from non-pupal mating clades were similar to the pupal mating clades. Furthermore, pupal mating species (which feed on the subgenera *Astrophea* and *Decaloba*) have a more similar *CNglcs* profile among them in contrast to non-pupal mating species (which feed on the subgenus *Passiflora*; Figure 3.6A). Consistently, both axes of the NMDS revealed significantly differences between these two groups (Figure 3.6A). The majority of individuals from the silvaniforms, melpomene, wallacei, and doris clades were primarily synthesizers. In contrast, just over half of the individuals from the *sara/sapho* and erato clades were primarily catchers, while the rest were synthesizers. Only a handful of individuals from different species were equally synthesizers and catchers (Figure 3.6B). Significant differences were only found in NMD1 between synthesizers and catchers, as well as between synthesizers and individuals that performed both process equally (Figure 3.6B). Finally, we found a

small but significant association between mimetic rings with the CNglc profile (Figure S3.4).

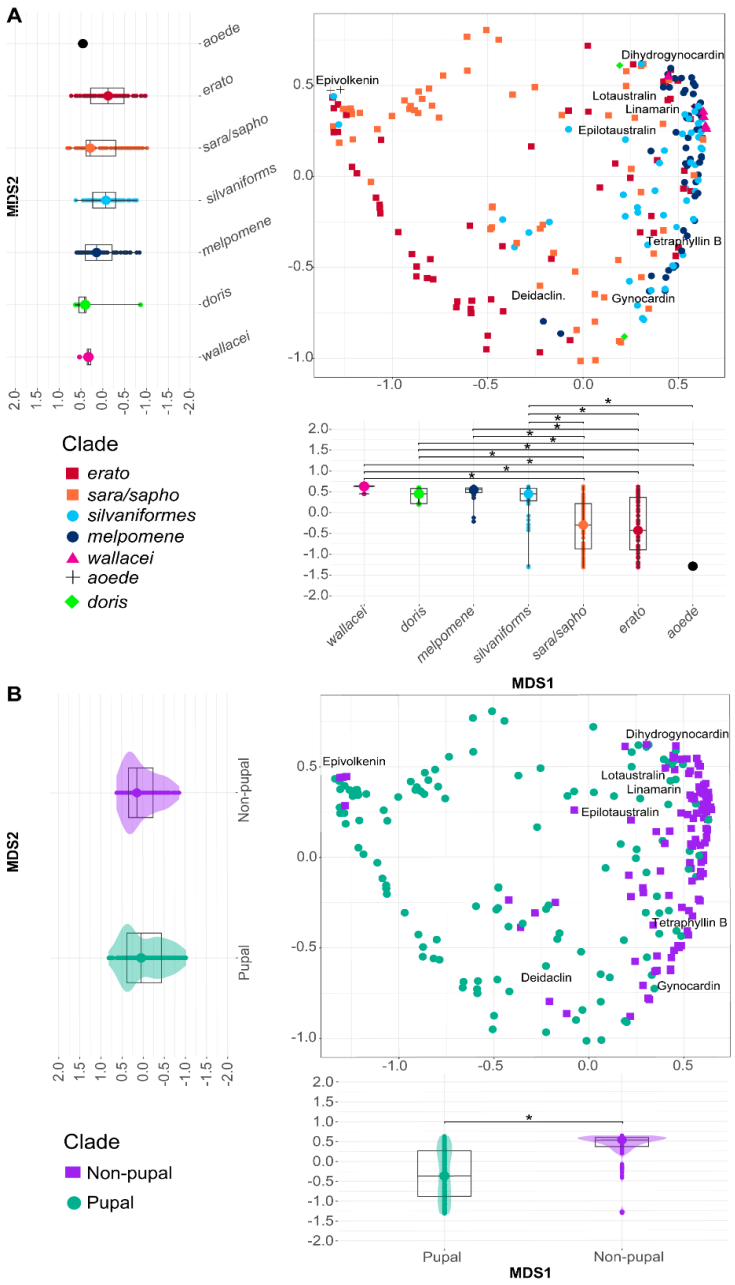

Figure 3.5 Non-metric multidimensional scaling (NMDS) of cyanogenic glycosides in *Heliconius* comparing between different clades. (A) Comparison between phylogenetic clades, and (B) comparison between pupal and non-pupal mating clades. The distance between two symbols is indicative of how different two individuals are in the composition of CNglcs. Boxplots for both NMDS1 and NMDS2 show differences between species in each axis. Asterisk (*) symbolises a p-value < 0.05.

Discussion

Our findings suggest that variation in the presence and concentration of cyanogenic glucosides in individuals of *Heliconius* is mostly due to the ability of a species to synthesize or sequester these compounds, phylogenetic relationships and host plant specialization. In contrast, variation in cyanogenic glucosides in these butterflies is less explained by the mimicry ring to which a species belongs. While the effect of phylogeny on the presence and concentration of cyanogenic glucosides in *Heliconius* has been previously reported (28,31), it was due to species of *sara/sapho* having a unique cyanogenic profile defined by the exclusive presence of the sequestered epivolkenin and deidaclin. Contrary to these reports, we detected the presence of *de novo* synthesised cyanogenic glucosides in the *sara/sapho* clade, and our phylogenetic signal was the result of the clades *silvaniform*, *melpomene*, *wallacei* and *doris* having similar cyanogenic profiles, where the sequestered epivolkenin and deidaclin were also present but in small amounts. Therefore, the phylogenetic signal we observed is likely the result of the association between mating clades of *Heliconius* with host plants of subgenera of *Passiflora*, where pupal mating species are usually specialists and feed on *Astrophea* and *Decaloba* while non-pupal mating species are more generalist and feed on *Passiflora* (16,30).

These dietary restrictions in *Heliconius* has led to monophagous species, in particular *sara/sapho* clade species, virtually losing their ability to *de novo* synthesize cyanogen compounds making them entirely dependent on cyanogen sequestration from their host plant (28,30). However, here we found *de novo*-synthesized aliphatic cyanogens such as linamarin and lotaustralin in individuals of the *sara/sapho* clade. This finding has two possible explanations: either species of the *sara/sapho* clade did not lose their ability to *de novo* synthesise cyanogenic glucosides, or they sequester linamarin and lotaustralin from their host plant. The latter is supported by the fact that both compounds are known to be sequestered by Lepidoptera such as the moth *Z. filipendulae* (38) and the butterfly *H. melpomene* (39), and also by the fact that the host plants *Astrophea* and *Decaloba* contain linamarin and lotaustralin (28).

We did not find evidence of species that either performed biosynthesis only or sequestration only (28,30). In contrast, our results point to all species of *Heliconius* using both strategies, with 30% of individuals prioritising sequestration, 60% prioritising biosynthesis, and only 10% equally using both. This is indicative of a potential trade-off between biosynthesis and sequestration at the individual level but not at the species level. It was interesting to observe that *de novo* synthesis is a widely used strategy across species and individuals despite being metabolically costlier (28), which suggests that other factors, such as local host plant availability may be more decisive for an individual to determine which strategy to apply.

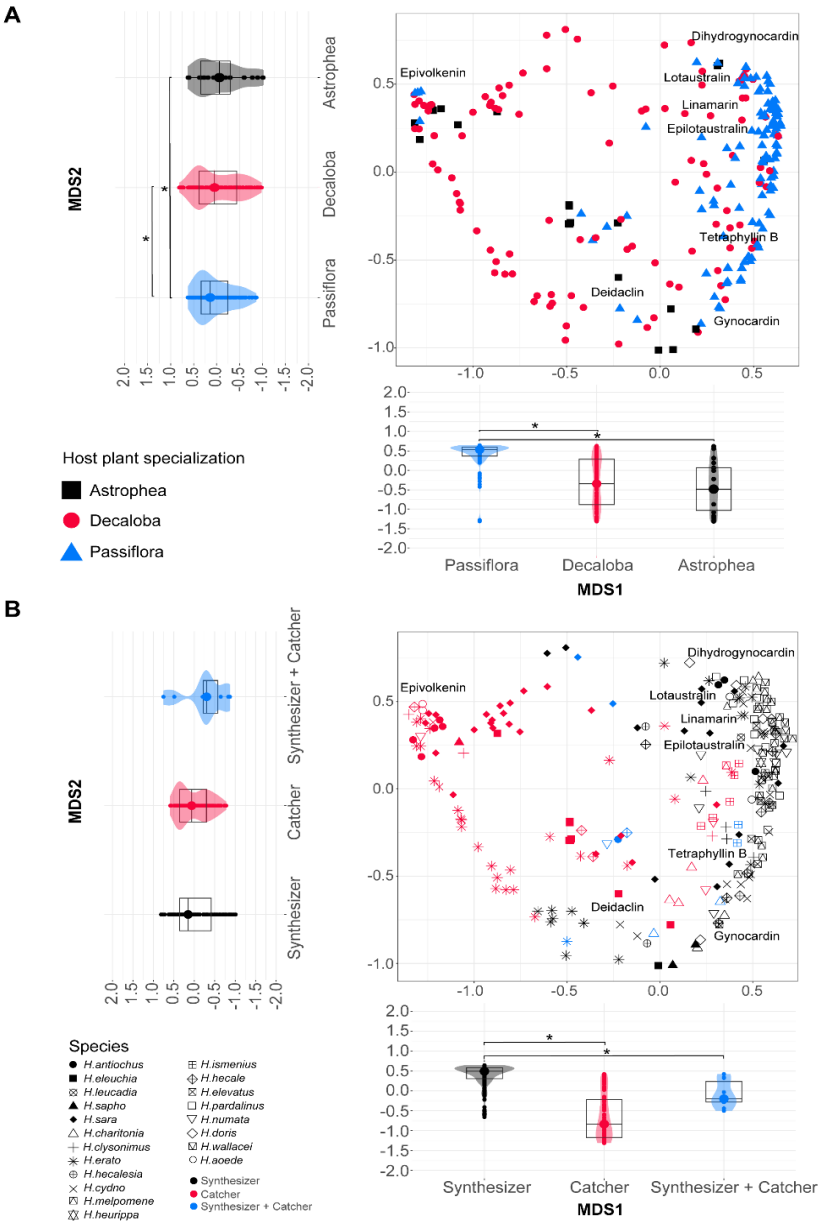

Figure 3.6 Non-metric multidimensional scaling (NMDS) of cyanogenic glycosides in *Heliconius* comparing between feeding and metabolic strategies. (A) Comparison based on host plant specialization, (B) comparison based on CNglcs acquisition. The distance between two symbols is indicative of how different two individuals are in the composition of CNglcs. Boxplots for both NMDS1 and NMDS2 show differences between individuals in each axis. Asterisk (*) symbolises a p-value < 0.05.

Although variation in cyanogenic profiles of *Heliconius* from different countries had been previously reported (31), here we found such variation at a finer geographic scale, specifically between biogeographic regions *sensu* Morrone et al. 2022 (32).

For example, *H. sara* and *H. erato* are much more toxic in Choco-Darien (Pacific coast) than in Imerí (Amazonia), while *H. melpomene* is more toxic in Sabana than in Magdalena. We even detected these variations within a single biogeographic region and mimicry ring. For instance, although *H. eleuchia eleuchia* and *H. cydno cydnides* both participate in the white-yellow mimicry ring and occur in the Cauca province, the concentration of CNglcs drastically varies between populations that are <250Km apart. Overall, these long and short range geographic variations in CNglcs may be the result of differences in the availability of host and pollen plants across localities (13,20,40) as well as to local predator communities being differentially sensitive to detect toxicity (26). Furthermore, contrary to previous evidence that showed different mimicry rings differing in cyanogenic profiles (16,31), with for example 'blue/yellow' being more toxic than 'postman' (16), we did not observe this pattern for either individual or combined CNglcs. The most notable example is *H. cydno*, a species that participates in three mimicry rings across its distribution but does not vary in its cyanogenic profile. Similarly, although *H. melpomene* participates in four mimicry rings, individuals from different rings do not differ in the composition or concentration CNglcs. We believe our conclusions on the relation between cyanogenic profiles and mimicry rings are more reliable than those from previous studies (31) since they analysed phenotypically similar species as part of the same mimicry ring even if they were not strict mimetic pairs.

Our findings support the existence of mutualistic relationships between both mimics and sympatric *Heliconius* since we observed sympatric species having similar cyanogenic concentrations regardless of them being mimics or non-mimics. Therefore, local communities of *Heliconius* seem to equally contribute to associative learning of predators (41). The only mimetic pair that did not follow this pattern was *H. erato* and *H. melpomene*: in all localities where they occur, we found one species being more toxic than the other. However, experimental evidence from domestic birds indicate that these cyanogenic variations between mimics should not impede local predator learning as long as each species has a toxin content >2 μg/mg (26). In line with the latter, 96% of the individuals we tested in this study had cyanogenic concentrations above 2 μg/mg, which shuould be enough to ensure their individual

protection. Interestingly, species in the *sara/sapho* clade were consistenly found as the most toxic within their local communities, and although protection from predators does not increase with cyanogenic concentration, the high toxicity of this clade may confer these species some protection from toxin-resistant predators (42,43). Alternatively, the high cyanogenic content in *sara/sapho* clade species may be a byproduct of their monophagy in *Astrophea* and *Decaloba* as previously suggested (28,30).

In summary, variation in cyanogenic patterns in *Heliconius* is not associated with mimicry rings as the same wing colour pattern exhibits different cyanogenic concentrations in different localities. The observed variation is rather explained by phylogenetic relationships and ecological adaptations such as host plant specialization, diversity of hostplants locally available, availability of precursors for biosynthesis of cyanogenic compounds in pollen-source plants, as well as the local predator community. This agrees with recent modelling and meta-analyses that showed that increased toxicity of preys does not translate into increased predator learning or generation of mimetic diversity (26).

References

1. Nahrstedt A. Relationships between the Defense Systems of Plants and Insects. In: Romeo JT, Saunders JA, Barbosa P, editors. Phytochemical Diversity and Redundancy in Ecological Interactions [Internet]. Boston, MA: Springer US; 1996 [cited 2023 Jun 14]. p. 217–30. (Recent Advances in Phytochemistry). Available from: https://doi.org/10.1007/978-1-4899-1754-6_8

2. Zagrobelny M, Bak S, Thorn Ekstrøm C, Erik Olsen C, Lindberg Møller B. The cyanogenic glucoside composition of Zygaena filipendulae (Lepidoptera: Zygaenidae) as effected by feeding on wild-type and transgenic lotus populations with variable cyanogenic glucoside profiles. Insect Biochem Mol Biol. 2007 Jan 1;37(1):10–8.

3. Davis RH, Nahrstedt A. Biosynthesis of cyanogenic glucosides in butterflies and moths: Effective incorporation of 2-methylpropanenitrile and 2-

methylbutanenitrile into linamarin and lotaustralin by Zygaena and *Heliconius* species (Lepidoptera). Insect Biochem. 1987 Jan 1;17(5):689–93.

4. Zagrobelny M, De Castro ÉCP, Møller BL, Bak S. Cyanogenesis in Arthropods: From Chemical Warfare to Nuptial Gifts. Insects. 2018 Jun;9(2):51.

5. Opitz SEW, Müller C. Plant chemistry and insect sequestration. Chemoecology. 2009 Sep 1;19(3):117–54.

6. Mattila ALK, Jiggins CD, Opedal ØH, Montejo-Kovacevich G, Castro ÉCP de, McMillan WO, et al. Evolutionary and ecological processes influencing chemical defense variation in an aposematic and mimetic *Heliconius* butterfly. PeerJ. 2021 Jun 18;9:e11523.

7. Pentzold S, Zagrobelny M, Khakimov B, Engelsen SB, Clausen H, Petersen BL, et al. Lepidopteran defence droplets - a composite physical and chemical weapon against potential predators. Sci Rep. 2016 Mar 4;6(1):22407.

8. Ruxton GD, Allen WL, Sherratt TN, Speed MP. Avoiding Attack: The Evolutionary Ecology of Crypsis, Aposematism, and Mimicry. 2nd edn. Oxford: Oxford University Press.; 2018.

9. Müller F. Ituna and Thyridia; a remarkable case of mimicry in butterflies. Trans Entomol Soc Lond 1879 Xx-Xxix. 1879;

10. Joron M, Mallet JL. Diversity in mimicry: paradox or paradigm? Trends Ecol Evol. 1998 Nov 1;13(11):461–6.

11. Mallet J. Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry. Ecol Entomol. 2010;35(s1):90–104.

12. Kuo CY. Predator learning can resolve the paradox of local warning signal diversity [Internet]. bioRxiv; 2023 [cited 2023 May 28]. p. 2023.05.04.539348. Available from: https://www.biorxiv.org/content/10.1101/2023.05.04.539348v1

13. Speed MP, Ruxton GD, Mappes J, Sherratt TN. Why are defensive toxins so variable? An evolutionary perspective. Biol Rev Camb Philos Soc. 2012 Nov;87(4):874–84.

14. Goodger JQD, Capon RJ, Woodrow IE. Cyanogenic polymorphism in Eucalyptus polyanthemos Schauer subsp. vestita L. Johnson and K. Hill (Myrtaceae). Biochem Syst Ecol. 2002 Jul 1;30(7):617–30.

15. Eisner HE. Defense mechanisms of arthropods. XX. Quantitative assessment of hydrogen cyanide production in two species of millipedes. Psyche (Stuttg). 1967;74(2):107–17.

16. Arias M, Meichanetzoglou A, Elias M, Rosser N, de-Silva DL, Nay B, et al. Variation in cyanogenic compounds concentration within a *Heliconius* butterfly community: does mimicry explain everything? BMC Evol Biol. 2016 Dec 15;16(1):272.

17. Skelhorn J, Rowe C. Frequency-dependent taste-rejection by avian predation may select for defence chemical polymorphisms in aposematic prey. Biol Lett. 2005 Dec 22;1(4):500–3.

18. Yezerski A, Gilmor TP, Stevens L. Genetic Analysis of Benzoquinone Production in Tribolium confusum. J Chem Ecol. 2004 May 1;30(5):1035–44.

19. Fahey SJ, Garson MJ. Geographic variation of natural products of tropical nudibranch Asteronotus cespitosus. J Chem Ecol. 2002 Sep;28(9):1773–85.

20. Saporito RA, Donnelly MA, Garraffo HM, Spande TF, Daly JW. Geographic and Seasonal Variation in Alkaloid-Based Chemical Defenses of Dendrobates pumilio from Bocas del Toro, Panama. J Chem Ecol. 2006 Apr 1;32(4):795–814.

21. Camara MD. Predator Responses to Sequestered Plant Toxins in Buckeye Caterpillars: Are Tritrophic Interactions Locally Variable? J Chem Ecol. 1997 Sep 1;23(9):2093–106.

22. Hay-Roe MM, Nation J. Spectrum of Cyanide Toxicity and Allocation in *Heliconius erato* and Passiflora Host Plants. J Chem Ecol. 2007 Feb 1;33(2):319–29.

23. Reudler JH, Lindstedt C, Pakkanen H, Lehtinen I, Mappes J. Costs and benefits of plant allelochemicals in herbivore diet in a multi enemy world. Oecologia. 2015 Dec;179(4):1147–58.

24. Bowers M. The evolution of unpalatability and the cost of chemical defense in insects. In: Insect chemical ecology: an evolutionary approach. New York: Chapman & Hall; 1992. p. 216–44.

25. Smilanich AM, Dyer LA, Chambers JQ, Bowers MD. Immunological cost of chemical defence and the evolution of herbivore diet breadth. Ecol Lett. 2009;12(7):612–21.
26. Chouteau M, Dezeure J, Sherratt TN, Llaurens V, Joron M. Similar predator aversion for natural prey with diverse toxicity levels. Anim Behav. 2019 Jul 1;153:49–59.
27. Briolat ES, Burdfield-Steel ER, Paul SC, Rönkä KH, Seymoure BM, Stankowich T, et al. Diversity in warning coloration: selective paradox or the norm? Biol Rev Camb Philos Soc. 2019 Apr;94(2):388–414.
28. de Castro ÉCP, Zagrobelny M, Zurano JP, Cardoso MZ, Feyereisen R, Bak S. Sequestration and biosynthesis of cyanogenic glucosides in passion vine butterflies and consequences for the diversification of their host plants. Ecol Evol. 2019 May 13;9(9):5079–93.
29. Wray V, Davis RH, Nahrstedt A. Biosynthesis of Cyanogenic Glycosides in Butterflies and Moths: Incorporation of Valine and Isoleucine into Linamarin and Lotaustralin by Zygaena and *Heliconius* Species (Lepidoptera). Z Für Naturforschung C. 1983 Aug 1;38(7–8):583–8.
30. Engler-Chaouat HS, Gilbert LE. De novo synthesis vs. sequestration: negatively correlated metabolic traits and the evolution of host plant specialization in cyanogenic butterflies. J Chem Ecol. 2007 Jan;33(1):25–42.
31. Sculfort O, de Castro ECP, Kozak KM, Bak S, Elias M, Nay B, et al. Variation of chemical compounds in wild Heliconiini reveals ecological factors involved in the evolution of chemical defenses in mimetic butterflies. Ecol Evol. 2020;10(5):2677–94.
32. Morrone JJ, Escalante T, Rodríguez-Tapia G, Carmona A, Arana M, Mercado-Gómez JD. Biogeographic regionalization of the Neotropical region: New map and shapefile. An Acad Bras Ciênc. 2022 Jan 31;94:e20211167.
33. Thodberg S, Sørensen M, Bellucci M, Crocoll C, Bendtsen AK, Nelson DR, et al. A flavin-dependent monooxygenase catalyzes the initial step in cyanogenic glycoside synthesis in ferns. Commun Biol. 2020 Sep 11;3(1):1–11.

34. Patil I. Visualizations with statistical details: The ‘ggstatsplot’ approach. J Open Source Softw. 2021 May 25;6(61):3167.
35. Oksanen FJ. Vegan: Community Ecology Package. R package Version 2.4-3. 2017.
36. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R Programming Tools for Plotting Data [Internet]. 2022 [cited 2023 Jun 14]. Available from: https://cran.r-project.org/web/packages/gplots/index.html
37. Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, et al. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 2014 Aug 1;30(15):2216–8.
38. Fürstenberg-Hägg J, Zagrobelny M, Jørgensen K, Vogel H, Møller BL, Bak S. Chemical Defense Balanced by Sequestration and De Novo Biosynthesis in a Lepidopteran Specialist. PLOS ONE. 2014 Oct 9;9(10):e108745.
39. Pinheiro de Castro ÉC, Demirtas R, Orteu A, Olsen CE, Motawie MS, Zikan Cardoso M, et al. The dynamics of cyanide defences in the life cycle of an aposematic butterfly: Biosynthesis versus sequestration. Insect Biochem Mol Biol. 2020 Jan;116:103259.
40. Jeckel AM, Grant T, Saporito RA. Sequestered and Synthesized Chemical Defenses in the Poison Frog Melanophryniscus moreirae. J Chem Ecol. 2015 May 1;41(5):505–12.
41. Speed MP. Muellerian mimicry and the psychology of predation. Anim Behav. 1993 Mar 1;45(3):571–80.
42. Williams BL, Brodie ED. Coevolution of Deadly Toxins and Predator Resistance: Self-Assessment of Resistance by Garter Snakes Leads to Behavioral Rejection of Toxic Newt Prey. Herpetologica. 2003;59(2):155–63.
43. Trigo JR. Effects of pyrrolizidine alkaloids through different trophic levels. Phytochem Rev. 2011 Mar 1;10(1):83–98.

CONCLUSIONS

The study of how species distribute and diversify is a major topic in evolutionary biology. In this thesis I have presented evidence contributing to our understanding of the biotic and biotic drivers of animal diversification. This has been carried out using a combination of biogeographic, genomic, and chemical tools in *Heliconius* butterflies.

First, isothermality, precipitation and topography are the abiotic factors that most influence the distribution patterns, species richness, phylogenetic diversity, and phylogenetic endemism of *Heliconius*. Niche similarity, on the other hand, does not seem to promote hybridization in these butterflies meaning that other factors, such as genetic distance, may have a major role in this process (1). I confirmed the foothills of the Eastern Cordillera of Colombia and the Amazonas basin are the regions with higher species richness and phylogenetic diversity in *Heliconius* (2). I also found high phylogenetic endemism in five Neotropical regions, which harbour species with unique wing colour pattern, for example: *H. heurippa*, *H. godmani* and *H. nattereri*, among others.

Second, I found that chromosomal rearrangements such as fusions may be involved in speciation events in the *sara/sapho* clade and thus be one of the biotic factors promoting diversification in *Heliconius*. Although my data does not allow me to pinpoint the evolutionary role or consequences of these fusions, it is likely these rearrangements contributed to the high and atypical chromosomic diversity within this clade (3). The sex-autosome fusions I found add evidence to the growing number of studies documenting the formation of neo-sex chromosomes in several animal groups, especially Lepidoptera (4–6).

Third, variation in cyanogenic toxicity is not a biotic factor contributing to the diversification of *Heliconius*. My results are consistent with mutualism between all

sympatric species of *Heliconius* regardless of them being mimics or not. Therefore, my thesis adds evidence that supports that higher toxicity levels do not necessarily translate into a higher or faster predator learning, or evolution of local mimicry. Other predator characteristics such as ability to learn and remember the warning signal may be more important (7,8).

References

1. Mallet J, Beltrán M, Neukirchen W, Linares M. Natural hybridization in heliconiine butterflies: The species boundary as a continuum. BMC Evol Biol. 2007;7:1–16.
2. Rosser N, Phillimore A, Huertas B, Willmott K, Mallet J. Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. Biol J Linn Soc. 2012;105(March):479–97.
3. Brown KS, Emmel TC, Eliazar PJ, Suomalainen E. Evolutionary patterns in chromosome numbers in neotropical Lepidoptera. I. Chromosomes of the Heliconiini (family Nymphalidae: subfamily Nymphalinae). Hereditas. 1992;117(2):109–25.
4. Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, et al. Neo-sex chromosomes and adaptive potential in tortricid pests. Proc Natl Acad Sci U S A. 2013 Apr 23;110(17):6931–6.
5. Yoshido A, Marec F, Sahara K. Resolution of sex chromosome constitution by genomic in situ hybridization and fluorescence in situ hybridization with (TTAGG) n telomeric probe in some species of Lepidoptera. Chromosoma. 2005 Aug;114(3):193–202.
6. Lucek K. Evolutionary Mechanisms of Varying Chromosome Numbers in the Radiation of Erebia Butterflies. Genes. 2018 Mar 16;9(3):166.
7. Chouteau M, Dezeure J, Sherratt TN, Llaurens V, Joron M. Similar predator aversion for natural prey with diverse toxicity levels. Anim Behav. 2019 Jul 1;153:49–59.

8. Kuo CY. Predator learning can resolve the paradox of local warning signal diversity [Internet]. bioRxiv; 2023 [cited 2023 May 28]. p. 2023.05.04.539348. Available from: https://www.biorxiv.org/content/10.1101/2023.05.04.539348v1

FUTURE DIRECTIONS

Each chapter of my thesis has opened new research avenues in which I will continue working in the immediate future with the help of my colleagues.

First, the distribution and diversity of *Heliconius* may be shaped by the distribution and diversity of their *Passiflora* host plants. Although the co-evolution of these two groups has been deeply studied, this question has not been tested at the geographic level, i.e., we do not know yet whether species richness in both groups coincides geographically and whether the spatial distribution of a butterfly and its hostplant overlaps completely. Therefore, a master's student (Catalina Sánchez) is currently using data I collected during my PhD working under the supervision of Dr. Marianne Elias in the Institut de Systématique, Evolution, Biodiversité of Paris (co-supervised by Dr. Camilo Salazar, Dr. Andrea Paz and me). Catalina will model the distribution of the species of *Passiflora* in the Americas, and her data will be used along mine (Chapter I) to conduct a special correlation analysis between *Passiflora* and *Heliconius*.

Second, because data from Chapter II did not allow me to decipher whether autosomes 4, 9 and 14 fused with the Z or the W chromosome (or both), we just generated new HiC data for the species where I found the SA fusions. These data are currently being mapped to the reference genome of *H. charithonia* which is the only genome that has the W chromosome assembled (kindly provided Dr. Adriana Briscoe from UCI). These new data will allow me to find which of the two (Z or W) chromosomes was involved in the SA fusion. Also, Dr. Camilo Salazar, Dr. Carolina Pardo, Dr. Joana Meier and I are generating reference genomes for all species in the *sara/sapho* clade at the Sanger Institute (UK). These new data will allow me to confirm the chromosome number in each species, identify additional chromosomal rearrangements, and test whether these rearrangements have contributed to adaptation or speciation.

Third, and lastly, wing colour pattern and toxicity have been explored separately in *Heliconius*, but the relation of these two traits has not been simultaneously addressed from a quantitative perspective. During my PhD I supervised undergraduate students in the 'Evolutionary Genomics Research Incubator' that took reflectance measures in all patches of colour in the wings of all individuals included in the toxicity analyses conducted in Chapter III. In each individual we measured at least three points per colour patch in the forewing and the hindwing both dorsally and ventrally, and under UV and non-UV filters. I will use these data to test whether contrasts in colour and brightness vary geographically and, more importantly, to test whether they vary along with toxicity (i.e, honesty of the warning signal).

SUPPLEMENTARY INFORMATION

CHAPTER 1

Environmental drivers of diversification and hybridization in neotropical butterflies

Table S1.1 Data sources and number of records per source.

Source	Number of records
Instituto de Ciencias Naturales, Bogotá - Colombia	1726
Instituto Alexander Von Humboldt	916
Jean Francois Le Crom personal collection	636
Universidad Nacional de Colombia, Medellín - Colombia	84
Universidad del Rosario, Bogotá - Colombia*	5973
Universidad de los Andes, Bogotá - Colombia	1316
Rosser et al. 2012 database	58062
Mallet et al. 2007 (https://www.ucl.ac.uk/taxome/hyb/hybtab.html)	154
Massardo, M., et al, 2020; Zikán, M., et al, 2017; Brown, K., et al, 1973	10
Total	**68,877**

*2078 individuals collected for this study were deposited in this collection.

Table S1.2 Hybridization cases included in this study.

INTERSPECIFIC HYBRIDIZATION	
Hybridizing pair	**Number of records**
H. cydno cordula x *H. melpomene melpomene*	7
H. cydno cydno x *H. melpomene martinae*	18
H. cydno weymeri x *H. melpomene vulcanus*	8
H. cydno zelinde x *H. melpomene vulcanus*	9
H. cydno cydno x *H. melpomene ca rosina*	2
H. cydno hermogenes x *H. melpomene melpomene*	1
H. cydno gadouae x *H. melpomene melpomene*	1
H. cydno barinasensis x *H. melpomene melpomene*	1
H. cydno galanthus x *H. melpomene rosina*	2
H. cydno galanthus x *H. pachinus*	3
H. cydno chioneus x *H. melpomene rosina*	1
H. hecale vetustus x *H. melpomene melpomene/thelxiopeia*	1
H. hecale fortunatus x *H. melpomene melpomene*	1
H. hecale versicolor x *H. elevatus pseudocupidineus*	1
H. hecale zeus x *H. elevatus perchlorus*	1
H. ethilla metalilis x *H. melpomene melpomene*	3
H. ethilla latona x *H. numata euphone*	1
H. ethilla metalilis x *H. numata peeblesi*	1
H. ethilla narcaea x H. numata ethra	1
H. ethilla narcaea x H. besckei	5
H. heurippa x *H. melpomene melpomene*	4
H. himera x *H. erato cyrbia*	42
H. himera x *H. erato favorinus*	4
H. himera x *H. erato lativitta*	1
H. melpomene cythera x *H. cydno alithea*	6
H. numata superioris x *H. melpomene melpomene*	1
H. numata superioris x *H. melpomene meriana*	2
H. numata aurora x *H. melpomene malleti*	1
H. numata numata x *H. melpomene thelxipea*	2
H. timareta florencia x *H. melpomene malleti*	12
H. timareta linaresi x *H. timareta florencia*	15
H. timareta linaresi x *H. heurippa*	1
H. elevatus elevatus x *H. numata superioris*	1
H. erato petiveranus x *H. charitonia vasquezae*	1

H. charitonia x H. peruvianus	1
H. hecalesia octavia x H. hortense	1
H. hecalesia formosus x H. clysonimus montanus	1
TOTAL INTERSPECIFIC HYBRIDS	**164**

INTRASPECIFIC HYBRIDIZATION	
Hybridizing pair	**Number of records**
H. clysonimus clysonimus x *H. cly. hygiana*	13
H. cydno chioneus x *H. cydno zelinde*	5
H. cydno chioneus x *H. cydno weymeri*	18
H. cydno weymeri x *H. cydno zelinde*	19
H. cydno zelinde x *H. cydno cydnides*	17
H. cydno cordula x *H. cydno wanningeri*	1
H. cydno weymeri x *H. cydno cydnides*	466
H. cydno cydno x *H. cydno hermogenes*	1
H. erato chestertonii x *H. erato venus*	25
H. erato colombina x *H. erato guarica*	20
H. erato dignus x *H. erato lativitta*	5
H. erato hydara x *H. erato lativitta*	37
H. erato reductimacula x *H. erato lativitta*	30
H. erato erato x *H. erato hydara*	9
H. erato venustus x *H. erato phyllis*	6
H. erato microlea x *H. erato luscombei*	17
H. erato notabilis x H. erato lativitta	9
H. erato favorinus x H. erato emma	4
H. erato hydara x H. erato demophon	19
H. hecale holcophorus x *H. h. melicerta*	23
H. ismenius boulleti x *H. i. metaphorus*	1
H. melpomene ca rosina x *H. m. martinae*	10
H. melpomene melpomene x *H. m. malleti*	36
H. melpomene penelope x *H. m. amaryllis*	10
H. melpomene belulla x *H. m. malleti*	4
H. melpomene ca rosina x *H. m. vulcanus*	1
H. melpomene pleseni x H. m. malleti	9
H. melpomene rosina x H. m. melpomene	20
H. m. vulcanus x H. m. melpomene	2
H. m. vulcanus x H. m. rosina	2
H. numata euphone x *H. numata aurora*	4

H. numata euphone x *H. numata silvana*	2
H. elevatus elevatus x *H. e. taracuanus*	2
H. eleuchia eleuchia x *H. eleuchia elusinus*	1
TOTAL INTRASPECIFIC HYBRIDS	**848**
TOTAL HYBRIDS	**1012**

Table S1.3 Species records used in this study, indicating the remaining data after filtering.

Species	Total records	Georeferenced records	Records with uncertainty >1km	Records after pruning
H. antiochus	963	722	5	181
H. aoede	1745	1308	2	354
H. astraea	119	107	0	91
H. atthis	356	320	0	59
H. besckei	438	394	0	110
H. burneyi	1458	1312	11	195
H. charitonia	4254	3190	13	769
H. clysonimus	1008	907	10	251
H. congener	172	154	7	55
H. cydno	3750	2812	31	352
H. demeter	556	500	3	80
H. doris	3092	2319	14	497
H. egeria	493	443	2	84
H. eleuchia	851	765	20	156
H. elevatus	954	715	8	201
H. erato	12074	9634	186	2057
H. eratosignis	379	341	0	52
H. ethilla	1446	1084	10	539
H. godmani	59	59	1	42
H. hecale	3461	3029	23	668
H. hecalesia	1048	943	2	168
H. hecuba	528	475	19	73
H. hermathena	137	137	7	137

H. heurippa	226	203	1	190
H. hewitsoni	216	194	0	115
H. hierax	150	134	0	150
H. himera	178	160	2	177
H. hortense	447	402	2	132
H. ismenius	1686	1517	17	400
H. leucadia	301	270	1	89
H. luciana	27	27	0	45
H. melpomene	7672	5665	98	1162
H. metharme	286	257	1	286
H. nattereri	13	13	0	13
H. numata	7244	5442	45	876
H. pachinus	175	157	0	105
H. pardalinus	721	648	1	155
H. peruvianus	113	101	0	42
H. ricini	111	99	0	81
H. sapho	1125	1012	11	228
H. sara	3138	2353	27	774
H. telesiphe	668	601	1	134
H. timareta	779	770	2	742
H. tristero	24	21	1	24
H. wallacei	2250	1800	1	320
H. xanthocles	974	876	18	260
Total species records	**67,865**	**54,392**	**603**	**13,671**

Table S1.4 Relative importance of predictors (environmental variables) of the distribution of each species.

Species	Altitude	Precipitation seasonality	Precipitation of coldest quarter	Isothermality	Min temperature of coldest month
H. ethilla	10.42	22.33	11.29	47.91	8.05
H. melpomene	7.84	12.81	9.17	63.00	7.18
H. antiochus	10.19	16.37	8.29	53.96	11.20
H. doris	9.53	16.21	14.36	48.26	11.64
H. erato	7.49	17.83	14.23	48.55	11.90
H. sara	8.59	14.55	16.89	52.18	7.79
H. hecale	13.32	18.61	17.17	38.54	12.36
H. sapho	22.11	15.97	14.85	30.00	17.07
H. clysonimus	20.89	11.21	13.96	40.76	13.18
H. cydno	10.43	8.46	19.46	52.08	9.56
H. ismenius	15.52	9.83	6.72	45.15	22.77
H. eleuchia	8.26	7.93	12.74	62.47	8.60
H. hecalesia	20.14	12.83	20.63	31.01	15.39
H. hortense	11.81	20.48	14.65	36.95	16.11
H. hierax	39.80	11.78	15.63	17.22	15.57
H. telesiphe	43.00	31.99	9.20	9.96	5.85
H. congener	46.63	9.23	8.64	12.45	23.05
H. hecuba	51.47	8.35	9.30	12.78	18.11
H. himera	44.10	7.94	7.88	11.01	29.07
H. charitonia	18.56	20.79	20.00	22.50	18.15
H. pachinus	8.73	13.52	53.22	14.98	9.55
H. hewitsoni	7.12	9.54	62.16	13.95	7.23
H. godmani	13.45	17.09	40.36	16.23	12.87

H. leucadia	18.93	32.95	11.65	15.19	21.28
H. ricini	17.29	35.00	19.11	17.72	10.87
H. pardalinus	11.19	42.48	17.49	16.46	12.37
H. demeter	9.51	47.97	18.22	16.74	7.56
H. timareta	22.20	42.02	7.81	12.77	15.20
H. heurippa	15.13	47.40	9.06	13.96	14.45
H. burneyi	10.68	41.20	15.35	20.70	12.07
H. wallacei	10.80	47.11	12.59	15.44	14.05
H. aoede	8.99	39.60	16.85	14.41	20.15
H. elevatus	9.84	33.69	13.03	11.84	31.60
H. xanthocles	6.15	43.24	27.75	15.32	7.55
H. numata	8.73	46.98	11.54	22.48	10.27
H. metharme	10.11	36.89	15.47	28.56	8.96
H. eratosignis	13.97	32.84	11.32	30.99	10.87
H. hermathena	25.01	35.69	12.39	17.86	9.05
H. luciana	18.47	28.77	18.57	20.53	13.65
H. besckei	14.55	29.60	11.33	21.79	22.73
H. egeria	15.24	37.80	14.53	18.75	13.68
H. astraea	9.46	49.74	19.09	12.28	9.44
H. peruvianus	10.31	38.71	8.64	33.24	9.09
H. atthis	13.50	44.96	14.98	13.98	12.58

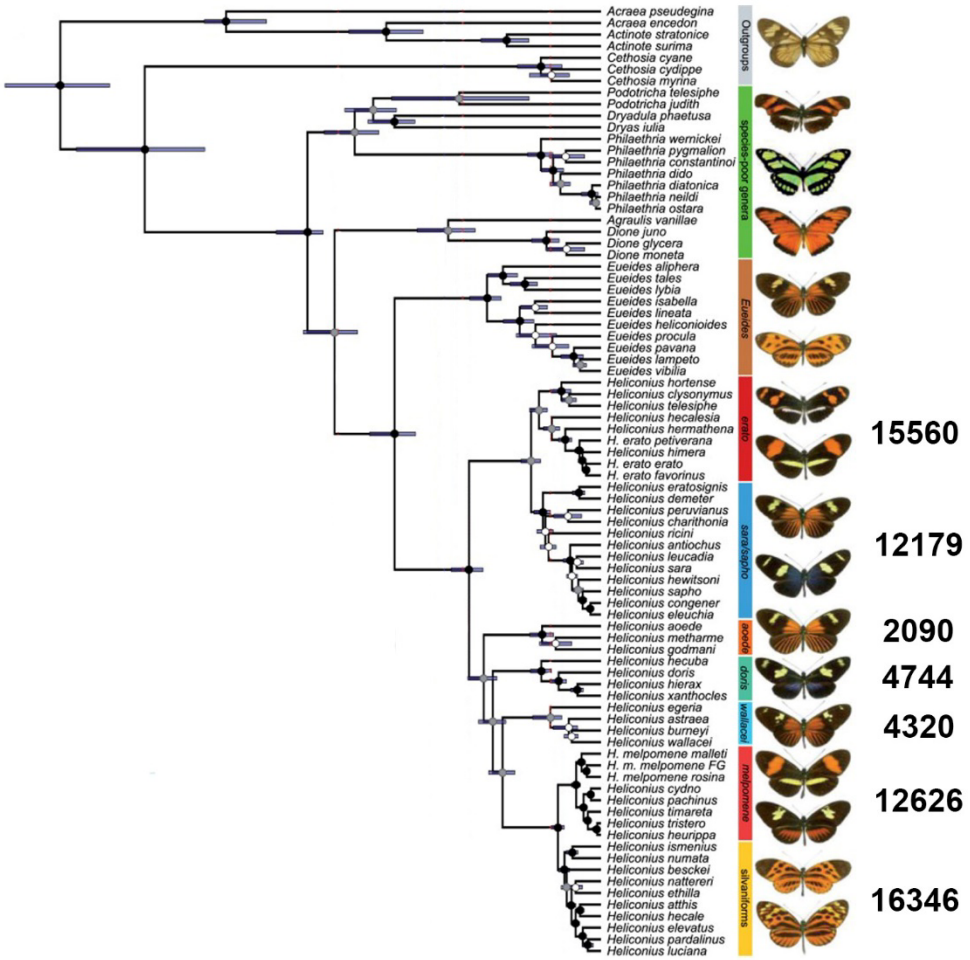

Figure S1.1 Records per clade across the phylogeny of *Heliconius* used in this study. Numbers indicate occurrence data per clade (which coincides with column 1 in Table S3). Hybrid records (1,012) are not included in this figure. Phylogeny modified from Kozak et al., (2015).

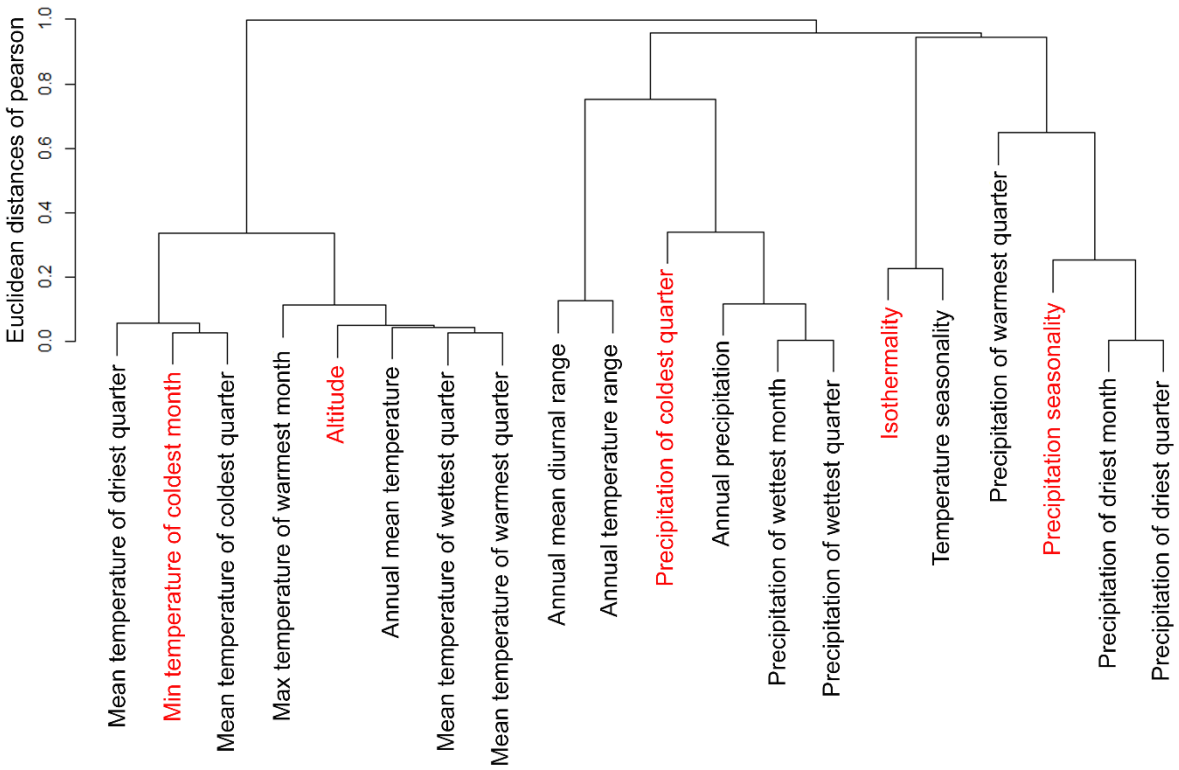

Figure S1.2 Selection of uncorrelated environmental variables.

Hierarchical clustering analysis using Pearson correlation values resulting from all possible pairwise comparisons between the initial 20 variables. Red: variables with VIF <5 (less collinear).

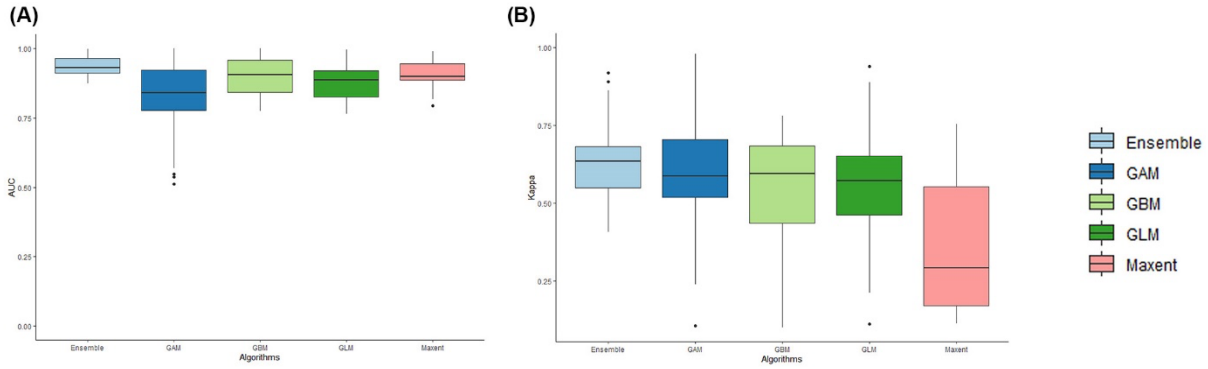

Figure S1.3 Measures of model performance. (A) AUC (B) Cohen's *kappa.*

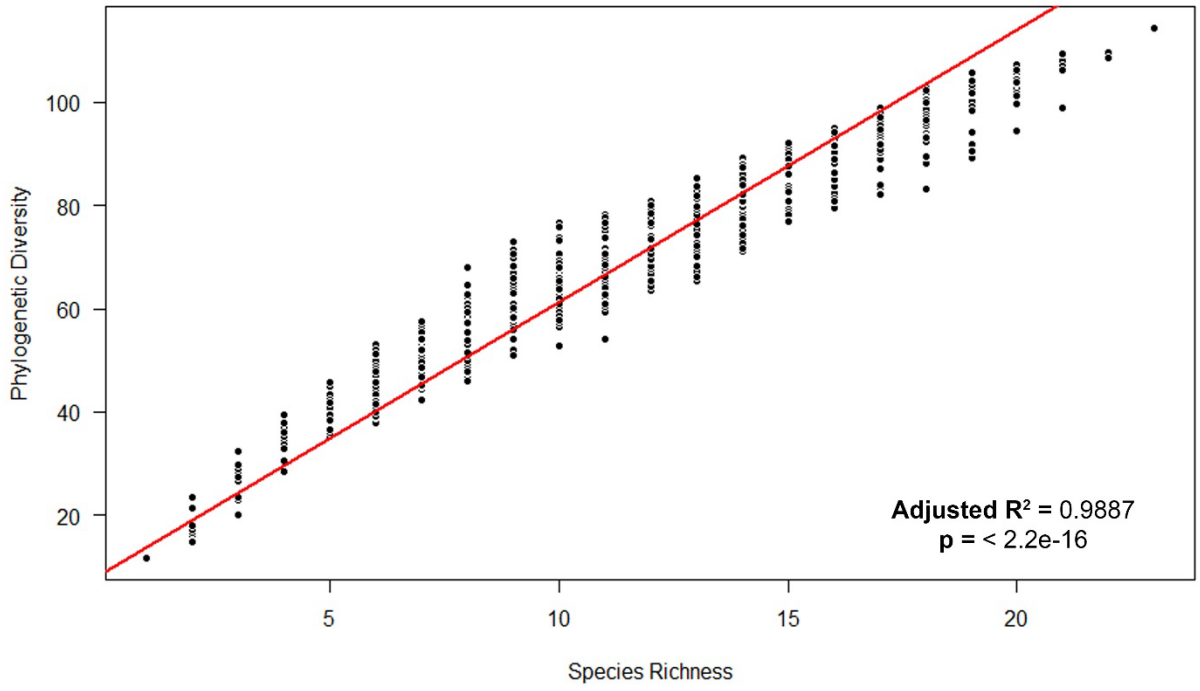

Figure S1.4 Linear regression model between phylogenetic diversity and species richness.

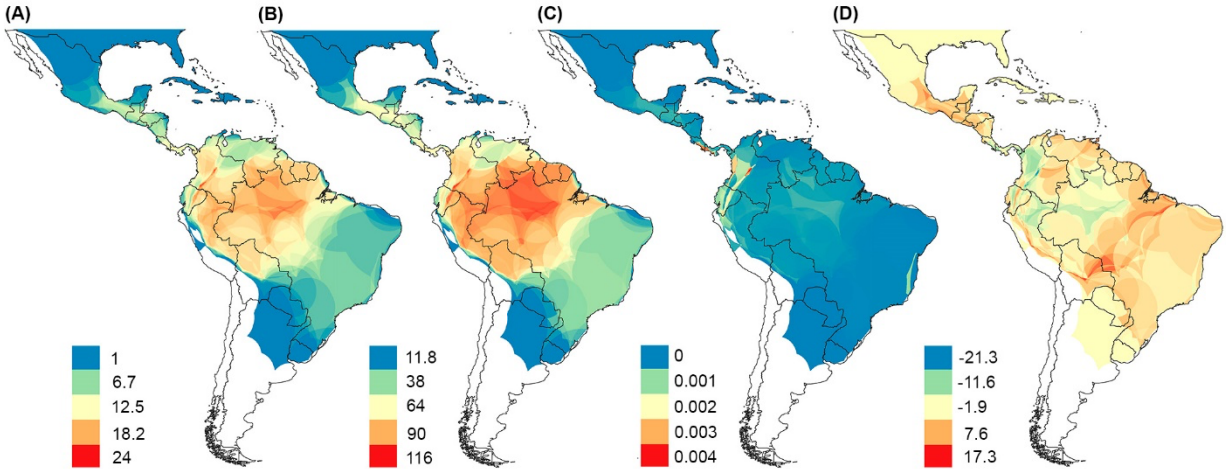

Figure S1.5 Maps of diversity metrics at 5 km. (a) species richness, (b) phylogenetic diversity, (c) phylogenetic endemism, and (d) residuals of phylogenetic diversity regressed on species richness. Warm colors indicate higher values, while cold colors are indicative of lower values.

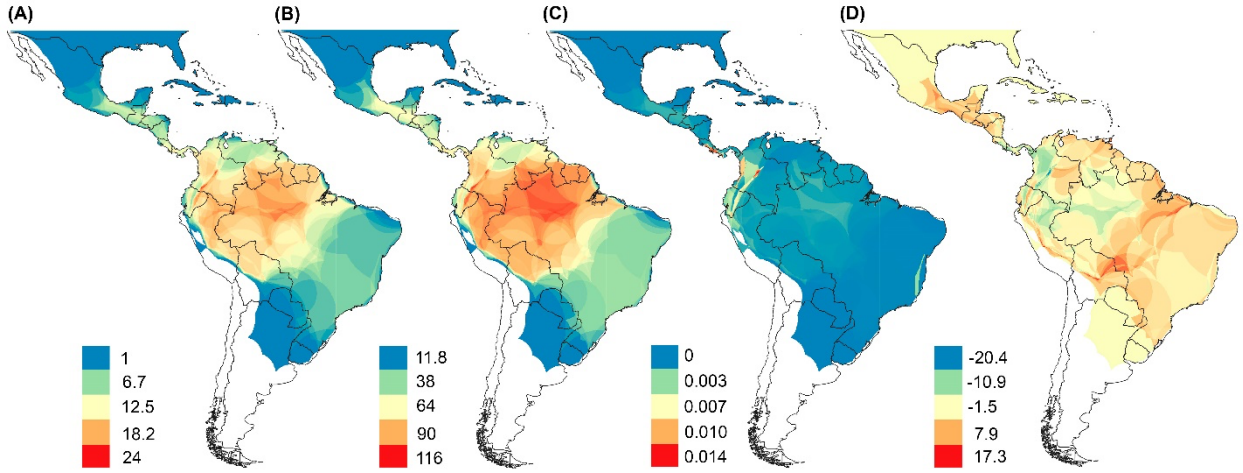

Figure S1.6 Maps of diversity metrics at 10 km. (a) species richness, (b) phylogenetic diversity, (c) phylogenetic endemism, and (d) residuals of phylogenetic diversity regressed on species richness. Warm colors indicate higher values, while cold colors are indicative of lower values.

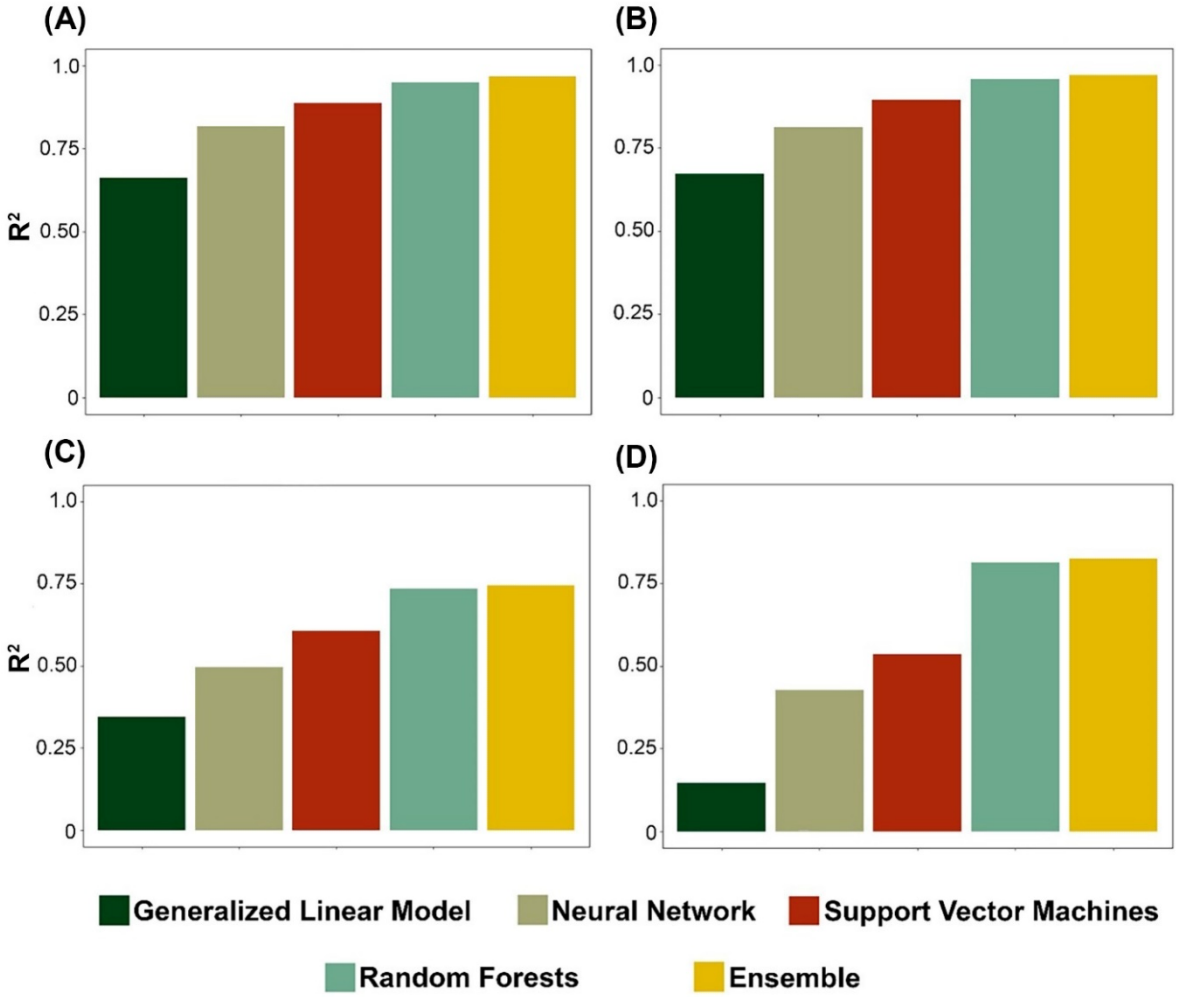

Figure S1.7 R^2 of algorithms and ensemble machine learning models for each measure of diversity. (a) Species richness, (b) phylogenetic diversity, (c) phylogenetic endemism, and (d) residuals of the phylogenetic diversity/species richness regression.

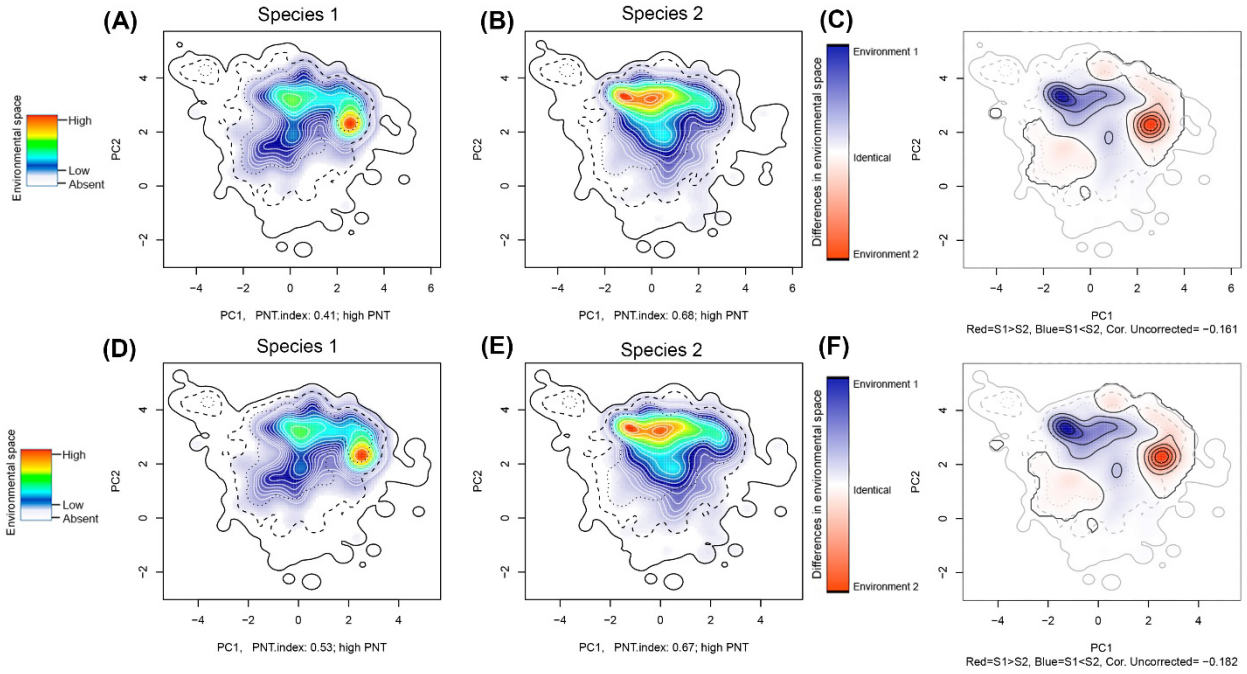

Figure S1.8 Assessment of niche similarity. Niche Overlap Test (NOT - top row) and Niche Divergence Test (NDT - bottom row) between *H. ethilla* (species 1) and *H. numata* (species 2). (a) and (d) environmental space of species 1; (b) and (e) environmental space of species 2; (c) and (f) difference in the environmental space (E-space) of two species and Niche E-space Correlation Index (NECI). When NECI was higher than 0.5, we corrected species occupied niches by the frequency of E-space in accessible environments. Significance of NOT and NDT can be found in Table 1. Equivalency statistic and niche background statistic for each NOT and NDT can be found in https://doi.org/10.5281/zenodo.5149294.

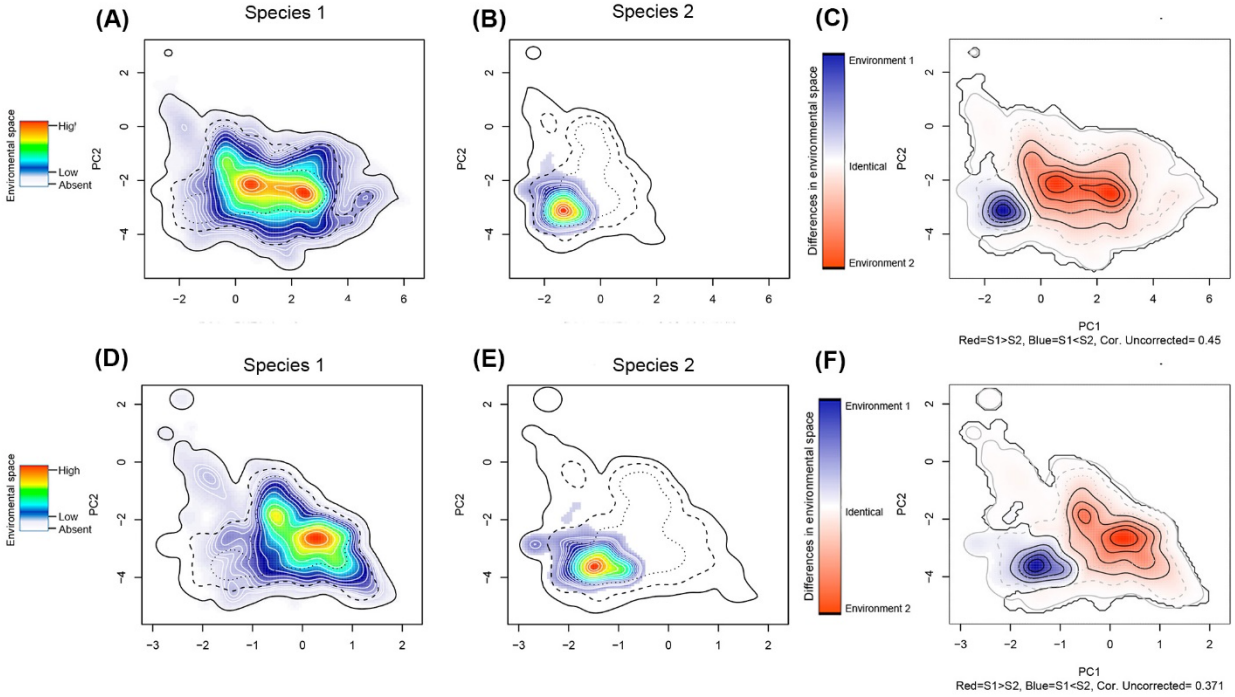

Figure S1.9 Assessment of niche similarity. Niche Overlap Test (NOT - top row) and Niche Divergence Test (NDT - bottom row) between *H. erato* (species 1) and *H. chestertonii* (species 2). (a) and (d) environmental space of species 1; (b) and (e) environmental space of species 2; (c) and (f) difference in the environmental space (E-space) of two species and Niche E-space Correlation Index (NECI). When NECI was higher than 0.5, we corrected species occupied niches by the frequency of E-space in accessible environments. Significance of NOT and NDT can be found in Table 1. Equivalency statistic and niche background statistic for each NOT and NDT can be found in https://doi.org/10.5281/zenodo.5149294.

CHAPTER 2

Three potential sex chromosome – autosome fusions in *Heliconius* butterflies

Table S2.1 Sample information and genotyping statistics. Individuals in bold were not included in the analyses due to low depth and missing data**.**

Sequence ID	taxon	sex	country	lat	lon	mean read depth	missing data per individual	high quality calls, % QUAL >30	% mapping genome	Accession	Publication
M4096	*H. antiochus antiochus*	Female	Colombia	-4.161	-69.975	18.009	0.133	94.31	95.86	ERS14696507	This study
M4097	*H. antiochus antiochus*	Male	Colombia	-4.161	-69.975	18.776	0.133	94.39	95.91	ERS14696506	This study
S21203	*H. antiochus antiochus*	Female	Brazil	4.736	-56.802	18.742	0.131	94.42	96.01	ERS14696505	This study
S21229	*H. antiochus antiochus*	Male	Brazil	4.736	-56.802	16.861	0.136	94.69	96.17	ERS14696504	This study
S21234	*H. antiochus antiochus*	Female	Brazil	4.736	-56.802	18.804	0.132	93.78	96.24	ERS14696503	This study
S337_1	*H. antiochus antiochus*	Male	Brazil	-13.487	61.026	18.409	0.134	92.75	96.10	ERS14696502	This study
S4897	*H. antiochus aranea*	Male	Colombia	2.583	-72.717	17.363	0.138	93.15	96.11	ERS14696501	This study
R5074	*H. antiochus aranea*	Female	Colombia	5.344	-72.422	13.198	0.148	93.35	96.94	ERS14696500	This study
S9556	*H. antiochus aranea*	Female	Panama	8.915	-78.393	16.928	0.141	93.36	95.90	ERS14696499	This study
D5208	*H. antiochus aranea*	Male	Venezuela	7.676	-72.233	12.081	0.159	92.23	95.49	ERS14696510	This study

D5221	*H. antiochus aranea*	Male	Venezuela	7.676	-72.233	12.470	0.154	92.05	94.72	ERS14696511	This study
D5213	*H. antiochus aranea*	Male	Venezuela	7.676	-72.233	15.565	0.150	92.00	95.60	ERS14696512	This study
D5252	***H. antiochus araneides***	**Male**	**Venezuela**	**8.111**	**-72.246**	**7.415**	**0.195**	**92.03**	**95.20**	ERS14696513	This study
D5253	*H. antiochus araneides*	Male	Venezuela	8.111	-72.246	13.244	0.159	91.67	95.52	ERS14696514	This study
D5259	*H. antiochus araneides*	Male	Venezuela	10.453	-67.632	14.314	0.151	92.30	95.56	ERS14696515	This study
D5270	*H. antiochus salvini*	Male	Venezuela	7.308	-61.472	13.431	0.150	91.73	95.34	ERS14696516	This study
D5273	*H. antiochus salvini*	Female	Venezuela	7.308	-61.472	11.922	0.151	92.28	96.39	ERS14696517	This study
D5276	*H. antiochus salvini*	Female	Venezuela	7.308	-61.472	13.028	0.149	92.02	95.40	ERS14696518	This study
A323	*H. antiochus subsp nov.*	Female	Colombia	5.617	-72.300	17.027	0.138	89.63	95.57	ERS14696519	This study
R1163	*H. antiochus subsp nov.*	Male	Colombia	2.114	-74.786	14.205	0.144	92.58	96.36	ERS14696520	This study
S4895	*H. antiochus subsp nov.*	Female	Colombia	2.576	-72.714	22.068	0.130	93.16	95.74	ERS14696521	This study
A28	*H. congener aquilionaris*	Male	Colombia	1.178	-76.665	14.054	0.146	90.42	94.60	ERS14696752	This study
A3740	*H. congener aquilionaris*	Male	Colombia	1.803	-75.655	21.775	0.135	91.12	97.06	ERS14696753	This study
A3837	*H. congener aquilionaris*	Male	Colombia	3.498	-74.024	18.668	0.141	91.20	97.09	ERS14696754	This study
A4278	*H. congener aquilionaris*	Male	Colombia	1.219	-76.685	16.667	0.141	90.20	93.84	ERS14696755	This study
R806	*H. congener congener*	Female	Ecuador	-1.435	-78.419	13.592	0.149	91.83	97.74	ERS14696756	This study

R807	*H. congener congener*	Male	Ecuador	-1.435	-78.419	12.981	0.152	92.35	97.93	ERS14696757	This study
R808	*H. congener congener*	Male	Ecuador	-1.435	-78.419	12.481	0.153	92.16	97.69	ERS14696758	This study
R831	*H. congener congener*	Female	Ecuador	-4.045	-78.584	10.211	0.156	89.70	93.03	ERS14696759	This study
R843	***H. congener congener***	**Male**	**Ecuador**	**-4.144**	**-79.253**	**9.079**	**0.395**	**93.17**	**90.30**	ERS14696760	This study
S40203	*H. congener congener*	Male	Ecuador	-1.070	-77.457	15.491	0.145	91.89	97.42	ERS14696761	This study
S857	*H. eleuchia eleuchia*	Female	Panama	8.856	-79.879	17.049	0.150	93.30	97.36	ERS14696762	This study
S2083	*H. eleuchia eleuchia*	Male	Colombia	7.032	-73.054	19.261	0.148	93.24	97.66	ERS14696763	This study
A2458	*H. eleuchia eleuchia*	Male	Colombia	5.079	-74.566	13.739	0.157	90.79	96.29	ERS14696764	This study
S3106	*H. eleuchia eleuchia*	Male	Colombia	4.398	-75.206	10.966	0.166	93.67	94.13	ERS14696765	This study
A3154	*H. eleuchia eleuchia*	Male	Colombia	3.684	-76.526	17.977	0.147	90.23	96.68	ERS14696766	This study
S534	*H. eleuchia elesinus*	Male	Colombia	3.900	-76.633	19.474	0.138	95.01	97.87	ERS14696767	This study
A3646	*H. eleuchia elesinus*	Male	Colombia	3.518	-76.757	15.840	0.146	90.60	97.15	ERS14696768	This study
A4749	*H. eleuchia elesinus*	Female	Colombia	3.576	-76.781	14.314	0.145	89.75	97.71	ERS14696769	This study
R4864	*H. eleuchia elesinus*	Female	Colombia	3.903	76.691	13.957	0.148	93.65	98.38	ERS14696770	This study
R4871	*H. eleuchia elesinus*	Male	Colombia	3.903	76.691	12.870	0.151	93.36	98.20	ERS14696771	This study
S8970	*H. eleuchia elesinus*	Female	Panama	7.549	-78.203	15.216	0.145	92.72	94.75	ERS14696772	This study

S2852	*H. eleuchia primularis*	Male	Ecuador	-2.443	-79.051	14.708	0.148	92.91	97.68	ERS14696773	This study
R869	*H. eleuchia primularis*	Male	Ecuador	-3.776	-79.821	11.054	0.159	92.82	97.90	ERS14696774	This study
R870	*H. eleuchia primularis*	Female	Ecuador	-3.776	-79.821	12.015	0.155	92.53	97.69	ERS14696775	This study
R887	*H. eleuchia primularis*	Female	Ecuador	-3.649	-79.766	10.494	0.162	92.45	97.78	ERS14696776	This study
R888	*H. eleuchia primularis*	Male	Ecuador	-3.654	-79.790	11.058	0.159	92.53	97.95	ERS14696777	This study
S40343	*H. eleuchia primularis*	Male	Ecuador	0.080	-78.888	14.571	0.149	92.85	97.83	ERS14696778	This study
D3	*H. hewitsoni*	Male	Panama	7.797	-80.753	14.934	0.160	92.26	96.90	ERS14696779	This study
D6	*H. hewitsoni*	Female	Panama	7.797	-80.753	12.049	0.165	92.40	98.04	ERS14696780	This study
D7	*H. hewitsoni*	Male	Panama	7.797	-80.753	12.208	0.165	92.63	97.78	ERS14696781	This study
S33	*H. leucadia pseudorhea*		Peru	-10.676	-75.123	16.996	0.102	94.12	97.11	ERS14696782	This study
A4016	*H. leucadia pseudorhea*	Female	Colombia	-4.133	-69.941	18.693	0.100	90.48	98.05	ERS14696783	This study
A2790	*H. sapho sapho*	Male	Colombia	6.850	-73.028	17.775	0.133	90.22	91.86	ERS14696784	This study
R2794	*H. sapho sapho*	Female	Colombia	6.850	-73.028	12.221	0.149	93.58	92.92	ERS14696785	This study
R4510	*H. sapho sapho*	Female	Colombia	7.058	-73.383	15.252	0.138	94.14	98.22	ERS14696786	This study
R5091	*H. sapho sapho*	Female	Colombia	5.840	74.839	14.137	0.141	93.86	98.26	ERS14696787	This study
R5092	*H. sapho sapho*	Female	Colombia	5.840	74.839	14.571	0.135	92.87	97.98	ERS14696788	This study
R5237	*H. sapho sapho*	Male	Panama	9.124	-79.488	12.956	0.145	92.91	91.76	ERS14696789	This study
R5314	*H. sapho sapho*	Female	Panama	9.124	-79.488	12.889	0.144	92.55	97.81	ERS14696790	This study
R5324	*H. sapho sapho*	Male	Panama	9.124	-79.488	13.869	0.142	92.09	97.71	ERS14696791	This study
A3497	*H. sapho chocoensis*	Male	Colombia	3.991	-76.227	14.402	0.141	90.64	98.03	ERS14696792	This study
A3501	*H. sapho chocoensis*	Male	Colombia	3.991	-76.227	17.346	0.136	90.55	97.42	ERS14696793	This study

R3510	*H. sapho chocoensis*	Female	Colombia	3.849	-77.259	15.723	0.138	93.77	98.17	ERS14696794	This study
R3511	*H. sapho chocoensis*	Male	Colombia	3.849	-77.259	13.073	0.145	93.60	98.42	ERS14696795	This study
A3491	*H. sapho chocoensis*	Female	Colombia	3.991	-76.227	13.492	0.145	90.83	97.46	ERS14696796	This study
A4534	*H. sara brevimaculata*	Female	Colombia	3.941	-77.365	24.550	0.039	90.63	96.14	ERS14696797	This study
A4538	*H. sara brevimaculata*	Male	Colombia	3.941	-77.365	23.836	0.040	91.57	96.94	ERS14696798	This study
A4678	*H. sara elektra*	Female	Colombia	3.322	-76.633	24.947	0.051	90.93	97.09	ERS14696799	This study
A4680	*H. sara elektra*	Male	Colombia	3.322	-76.633	17.223	0.057	90.84	98.82	ERS14696800	This study
A4683	*H. sara elektra*	Female	Colombia	3.322	-76.633	18.660	0.055	91.96	98.07	ERS14696801	This study
S4704	*H. sara elektra*	Male	Colombia	3.322	-76.633	23.015	0.053	95.22	98.76	ERS14696802	This study
S29	*H. sara magdalena*	Male	Panama	8.709	-79.908	19.599	0.029	94.31	98.68	ERS14696803	This study
R31	*H. sara magdalena*	Male	Panama	8.709	-79.908	16.223	0.046	93.68	98.81	ERS14696804	This study
R32	*H. sara magdalena*	Male	Panama	8.709	-79.908	16.791	0.045	93.15	98.69	ERS14696805	This study
S2172	*H. sara magdalena*	Male	Venezuela	7.799	-72.199	25.416	0.046	95.25	98.86	ERS14696806	This study
S3533	*H. sara magdalena*	Male	Colombia	5.571	-77.501	24.810	0.040	94.70	98.66	ERS14696807	This study
A3900	*H. sara magdalena*	Male	Colombia	6.369	-77.378	20.491	0.042	91.86	98.66	ERS14696808	This study
R4134	*H. sara magdalena*	Male	Colombia	6.446	-70.688	18.466	0.057	93.69	98.91	ERS14696809	This study
R4186	*H. sara magdalena*	Male	Colombia	5.744	-74.229	14.595	0.048	94.31	99.03	ERS14696810	This study
R4187	*H. sara magdalena*	Male	Colombia	5.744	-74.229	18.422	0.044	93.08	98.74	ERS14696811	This study

A4340	*H. sara magdalena*	Male	Colombia	6.098	-67.486	17.671	0.058	91.87	98.34	ERS14696812	This study
S4346	*H. sara magdalena*	Female	Colombia	6.098	-67.486	17.273	0.059	95.14	98.54	ERS14696813	This study
D5248	*H. sara magdalena*	Male	Venezuela	8.069	-72.248	14.001	0.050	91.75	96.50	ERS14696814	This study
D5211	*H. sara magdalena*	Male	Venezuela	7.676	-72.233	13.666	0.056	92.66	98.64	ERS14696815	This study
S1	*H. sara sara*	Male	Brazil	-1.468	-48.440	19.233	0.058	93.48	87.39	ERS14696816	This study
S5	*H. sara sara*	Male	Peru	-10.653	-75.109	15.957	0.061	94.41	77.38	ERS14696817	This study
S8	*H. sara sara*	Male	Peru	-10.327	-74.943	20.504	0.056	93.97	97.85	ERS14696818	This study
S12	*H. sara sara*	Female	Brazil	-4.066	-54.847	19.816	0.057	92.66	97.97	ERS14696819	This study
R15	*H. sara sara*	Male	Brazil	-1.208	-54.800	14.194	0.063	93.10	98.66	ERS14696820	This study
R16	*H. sara sara*	Male	Brazil	-10.891	-55.440	15.055	0.062	93.33	98.54	ERS14696821	This study
R35	*H. sara sara*	Female	Peru	-5.288	-80.679	14.333	0.064	93.95	98.46	ERS14696822	This study
R36	*H. sara sara*	Female	Peru	-5.288	-80.679	16.818	0.045	91.84	97.98	ERS14696823	This study
R37	*H. sara sara*	Male	Peru	-5.288	-80.679	14.743	0.046	92.87	97.74	ERS14696824	This study
S1053	*H. sara sara*	Male	Colombia	4.175	-73.678	22.577	0.054	93.75	97.98	ERS14696825	This study
R812	*H. sara sara*	Female	Ecuador	-3.915	-79.740	14.352	0.062	92.79	98.67	ERS14696826	This study
R820	*H. sara sara*	Male	Ecuador	-2.868	-78.370	16.235	0.060	93.14	98.80	ERS14696827	This study
R828	*H. sara sara*	Male	Ecuador	-4.045	-78.584	15.138	0.061	92.79	98.73	ERS14696828	This study
R3467	*H. sara sara*	Male	Colombia	1.803	-75.655	14.345	0.061	93.72	99.14	ERS14696829	This study
R3751	*H. sara sara*	Male	Colombia	1.803	-75.655	15.775	0.060	93.41	98.77	ERS14696830	This study
A4121	*H. sara sara*	Female	Colombia	-3.770	-70.340	27.524	0.052	91.97	98.23	ERS14696831	This study
A4272	*H. sara sara*	Female	Colombia	1.215	-76.683	18.391	0.057	92.39	98.06	ERS14696832	This study
A4343	*H. sara sara*	Female	Colombia	6.098	-67.486	19.928	0.056	91.63	98.30	ERS14696833	This study
A4489	*H. sara sara*	Male	Ecuador	-1.037	-77.844	25.071	0.053	92.34	97.96	ERS14696834	This study
S21204	*H. sara sara*	Male	Surinam	4.736	-56.802	20.545	0.056	94.61	98.41	ERS14696835	This study
S40219	*H. sara sara*	Male	Ecuador	-1.037	-77.844	20.755	0.056	93.64	98.21	ERS14696836	This study
R871	*H. sara sprucei*	Male	Ecuador	-3.767	-79.825	15.049	0.047	92.73	97.60	ERS14696837	This study
R886	*H. sara sprucei*	Male	Ecuador	-3.653	-79.761	16.130	0.047	92.66	98.84	ERS14696838	This study
A3856	*H. sara sprucei*	Female	Colombia	2.964	-78.182	23.720	0.033	91.74	98.77	ERS14696839	This study

A3858	*H. sara sprucei*	Female	Colombia	2.966	-78.185	23.520	0.044	92.11	98.71	ERS14696840	This study
A3861	*H. sara sprucei*	Male	Colombia	2.964	-78.182	22.827	0.044	91.44	98.70	ERS14696841	This study
R4590	*H. sara sprucei*	Female	Colombia	3.941	-77.365	15.556	0.047	93.09	98.70	ERS14696842	This study
R4591	*H. sara sprucei*	Male	Colombia	3.941	-77.365	13.712	0.048	93.31	98.84	ERS14696843	This study
R4597	*H. sara sprucei*	Female	Colombia	3.941	-77.365	15.596	0.046	93.56	98.73	ERS14696844	This study
SRR4032025	*H. charitonia*		Puerto Rico	18.023	-66.829	16.567	0.062	91.60	79.23	SAMN05224120	Van Belleghem et al. 2017
SRR4032026	*H. charitonia*		Puerto Rico	18.023	-66.829	14.345	0.059	90.53	80.29	SAMN05224121	Van Belleghem et al. 2017
SAMEA53943 85	*H. sara*										This study

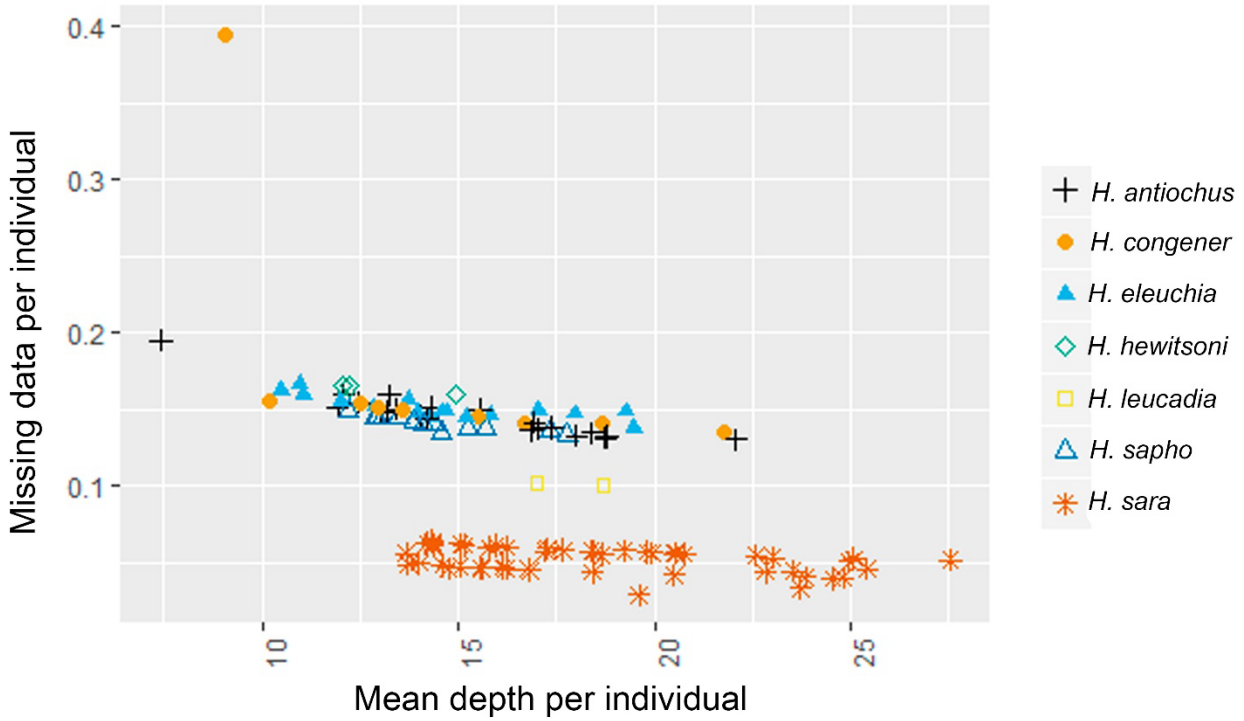

Figure S2.1 Missing data and mean depth per individual. Each species is symbolised by a unique symbol and colour. Note that the lower missing data proportion in *H. sara* is likely due to their closer similarity to the reference genome (*H. sara* female).

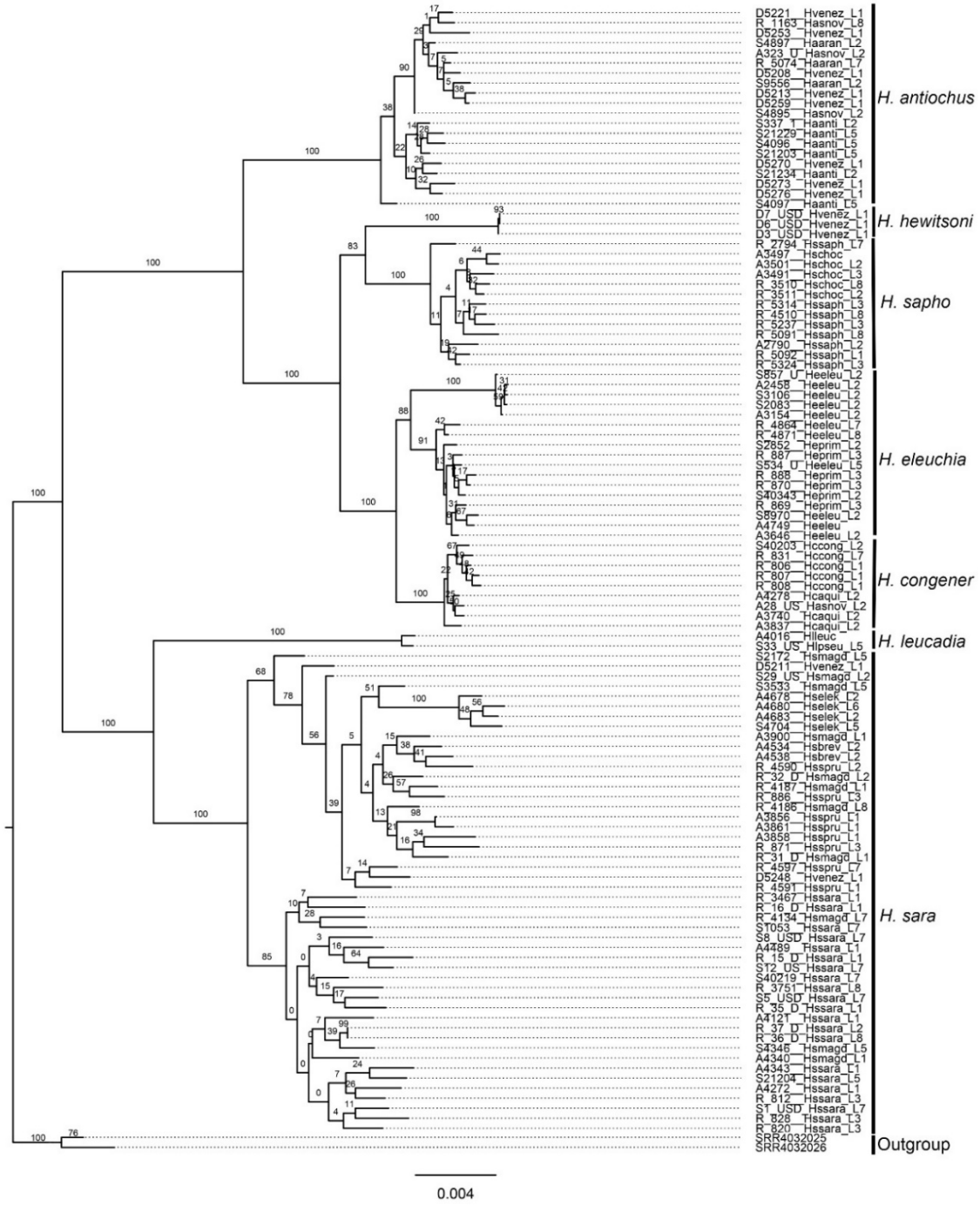

Figure S2.2 Maximum likelihood phylogeny of chromosome 1. Bootstrap support values are indicated at branches.

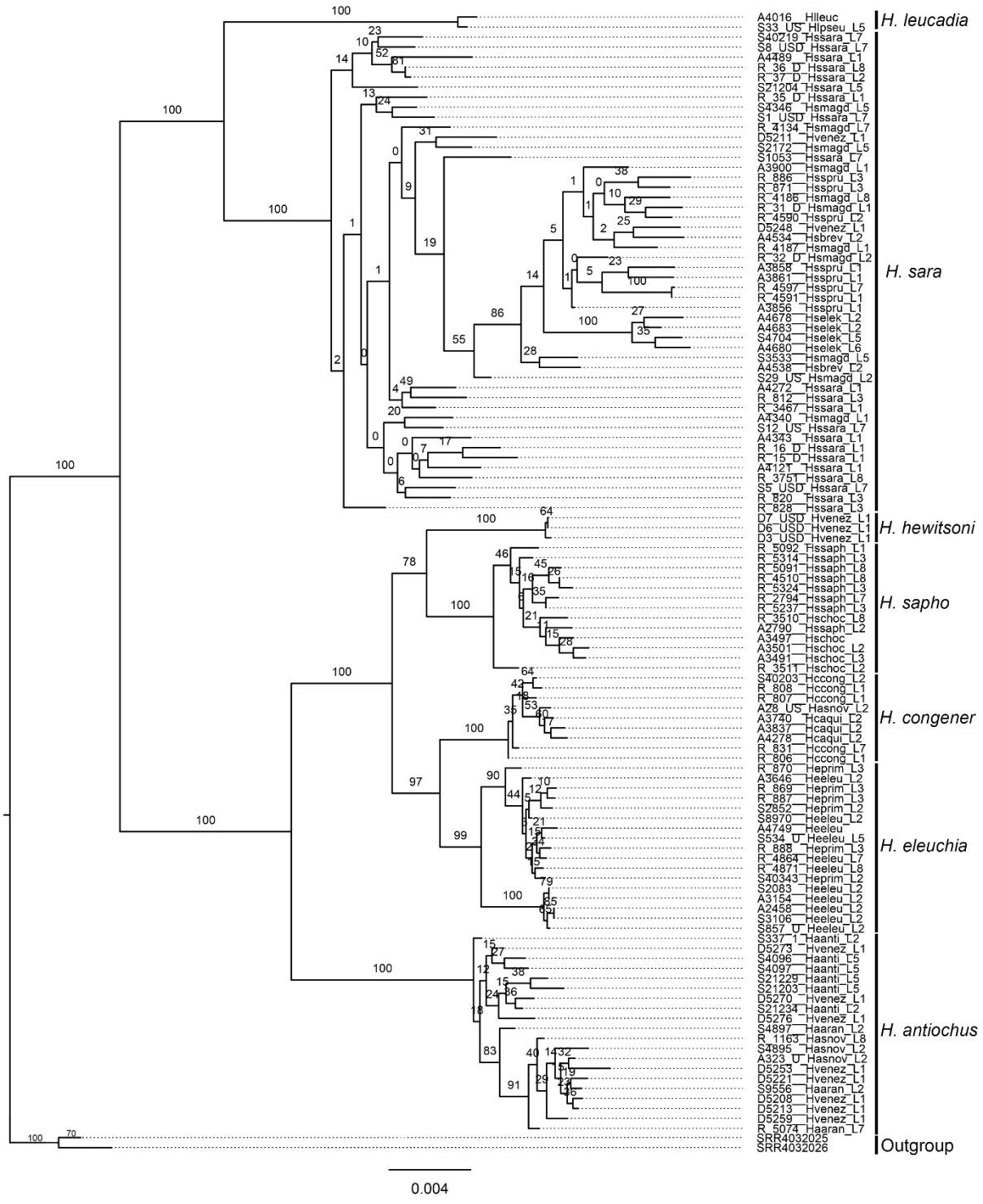

Figure S2.3 Maximum likelihood phylogeny of chromosome 2. Bootstrap support values are indicated at branches.

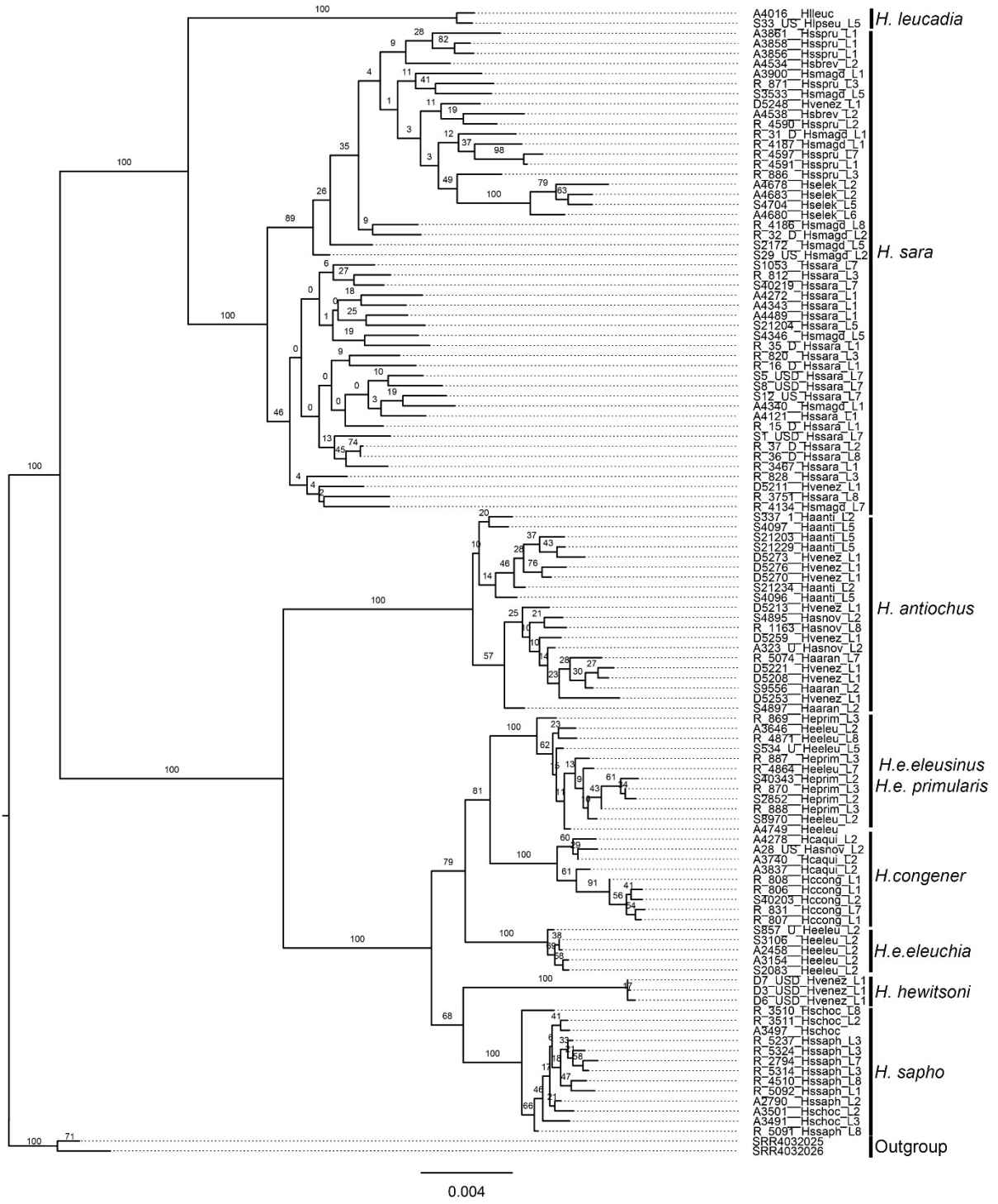

Figure S2.4 Maximum likelihood phylogeny of chromosome 3. Bootstrap support values are indicated at branches.

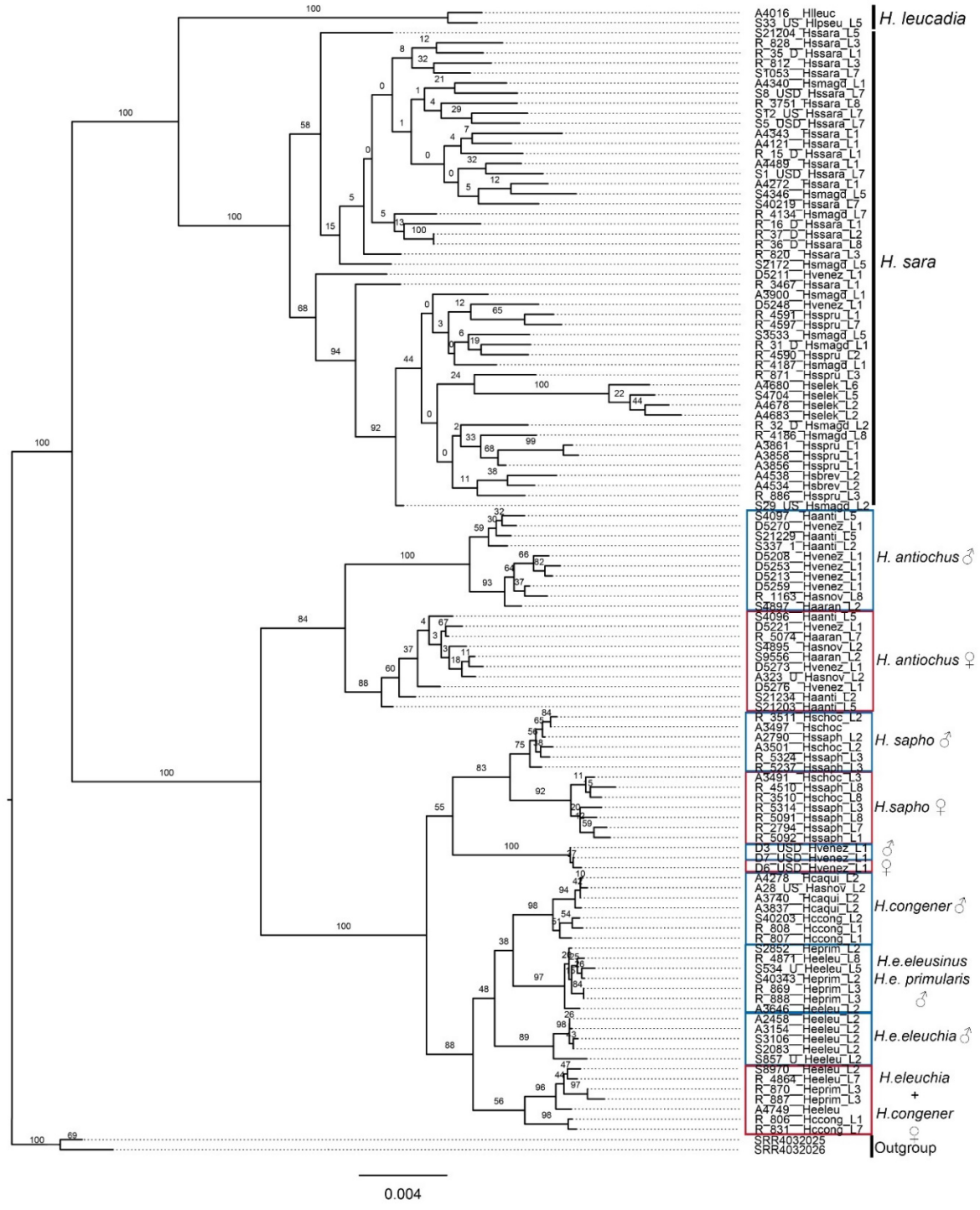

Figure S2.5 Maximum likelihood phylogeny of chromosome 4. Bootstrap support values are indicated at branches. Blue squares group males and while red squares group females.

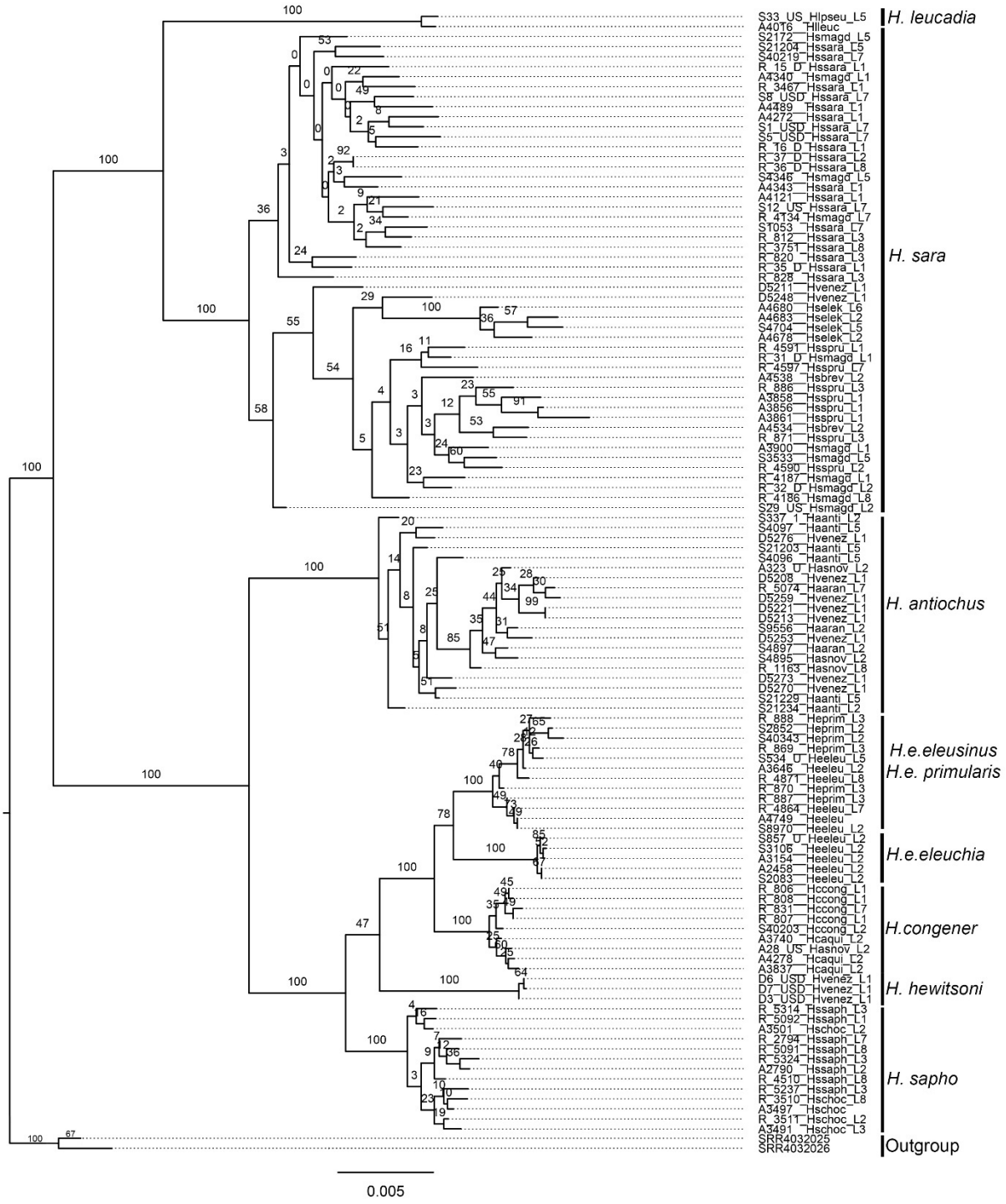

Figure S2.6 Maximum likelihood phylogeny of chromosome 5. Bootstrap support values are indicated at branches.

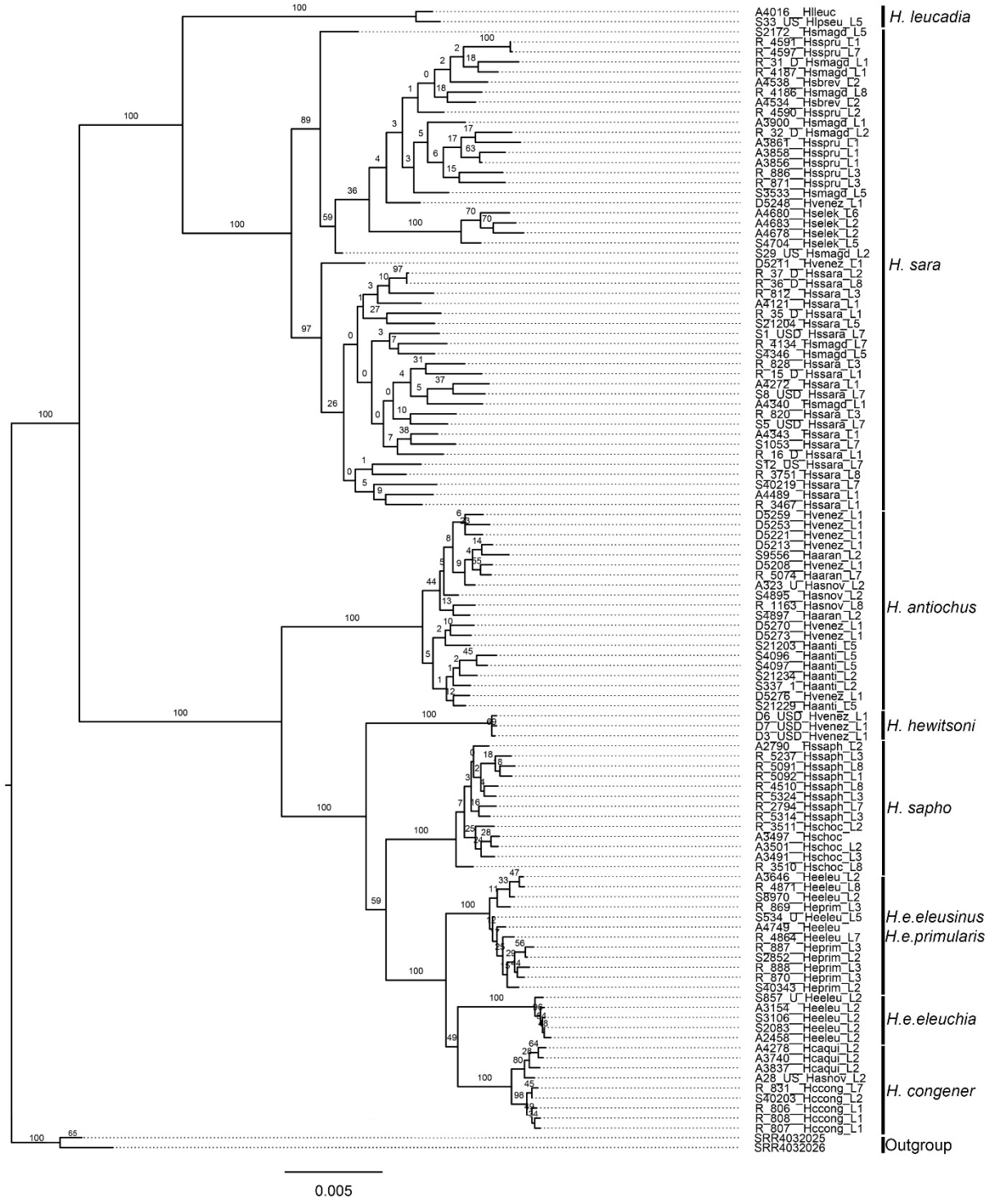

Figure S2.7 Maximum likelihood phylogeny of chromosome 6. Bootstrap support values are indicated at branches.

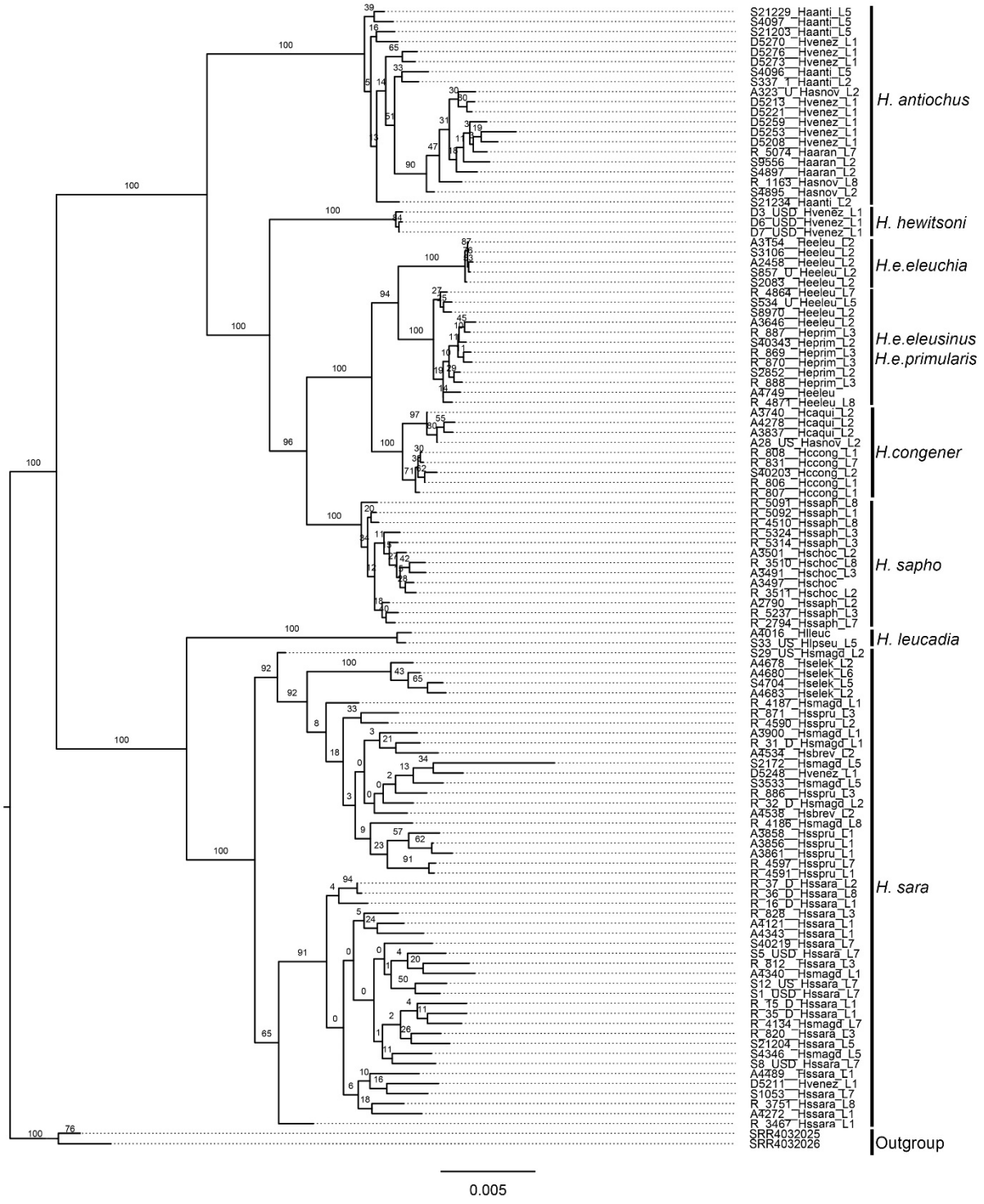

Figure S2.8 Maximum likelihood phylogeny of chromosome 7. Bootstrap support values are indicated at branches.

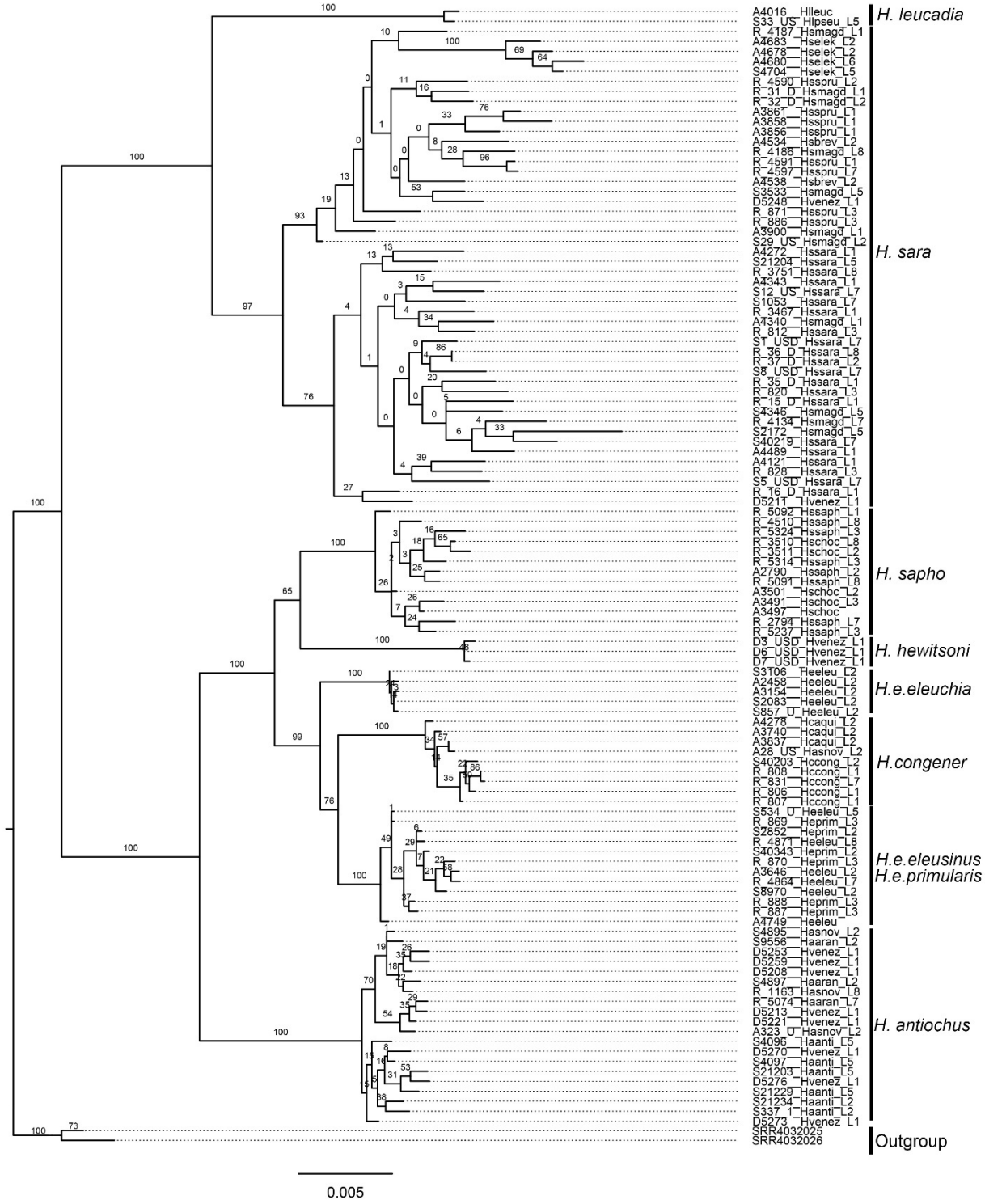

Figure S2.9 Maximum likelihood phylogeny of chromosome 8. Bootstrap support values are indicated at branches.

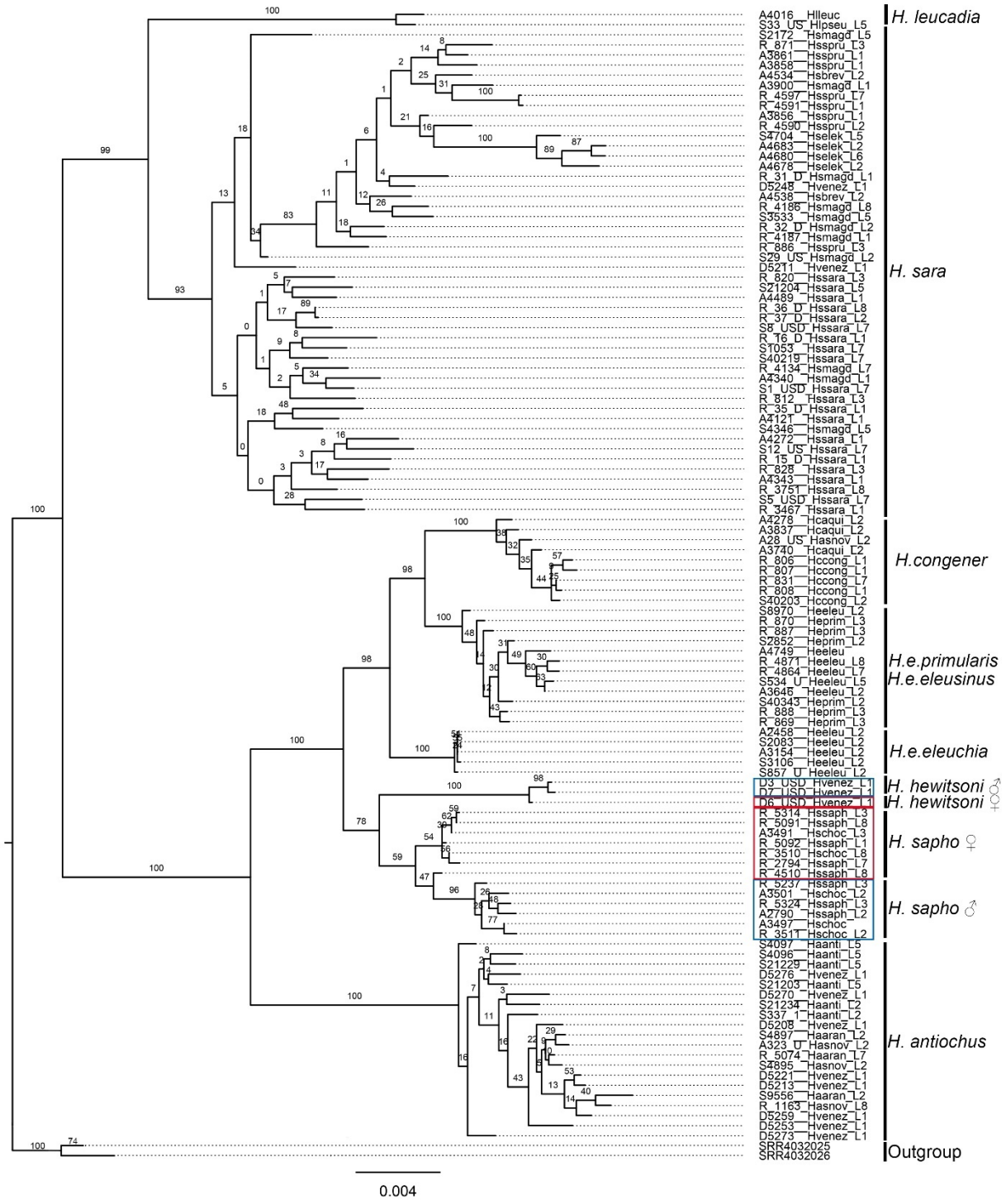

Figure S2.10 Maximum likelihood phylogeny of chromosome 9. Bootstrap support values are indicated at branches. Blue squares group males and while red squares group females.

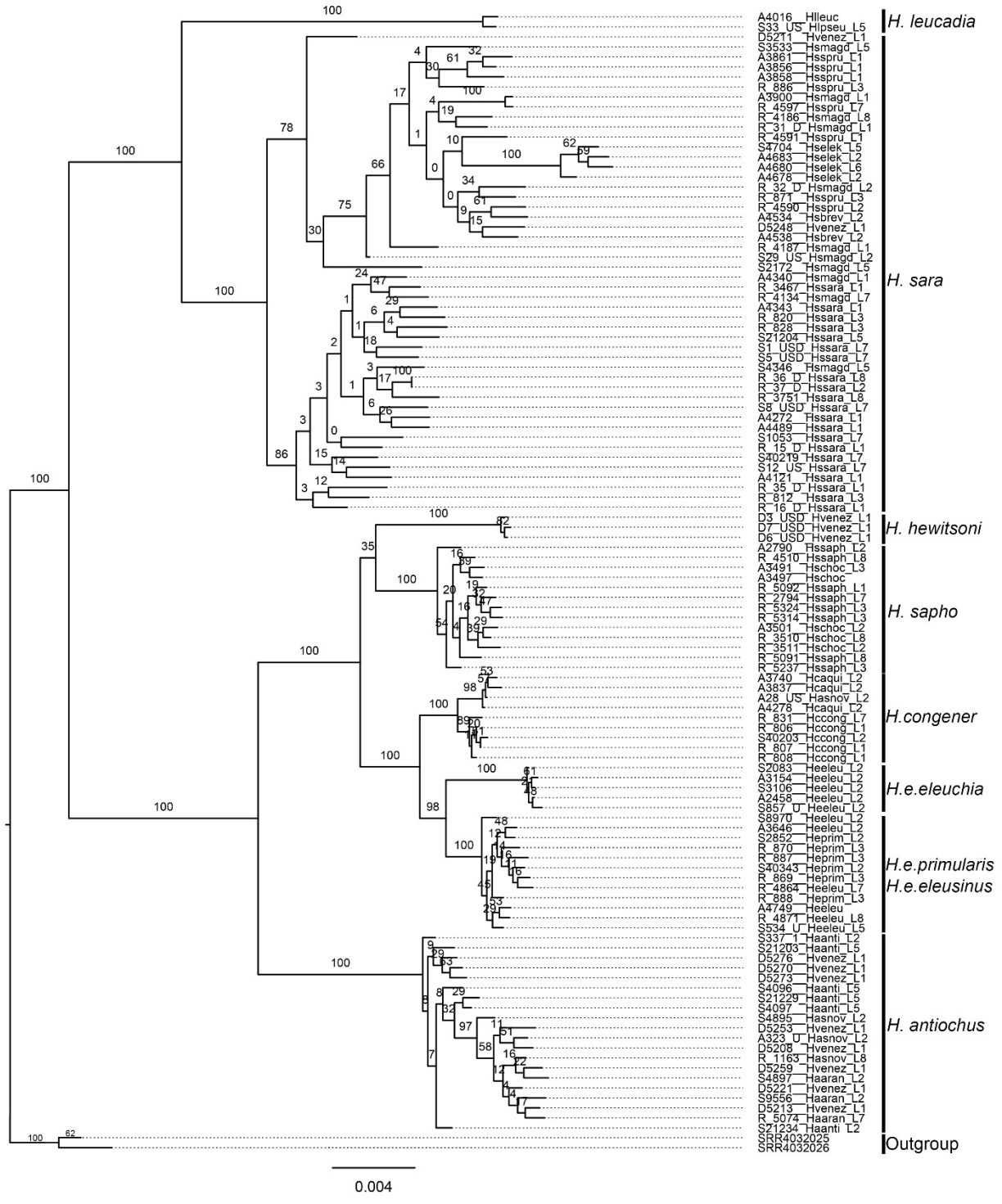

Figure S2.11 Maximum likelihood phylogeny of chromosome 10. Bootstrap support values are indicated at branches.

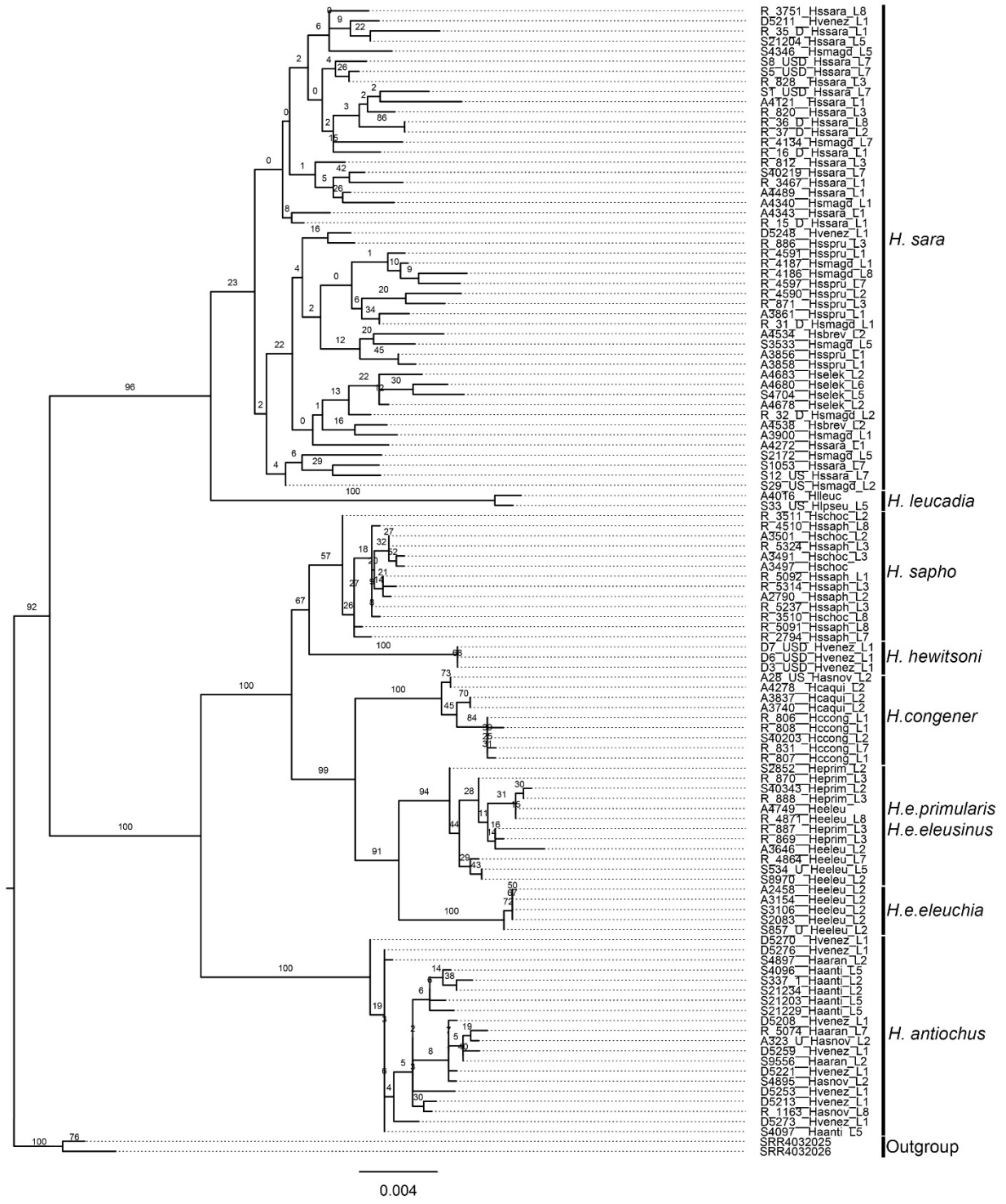

Figure S2.12 Maximum likelihood phylogeny of chromosome 11. Bootstrap support values are indicated at branches.

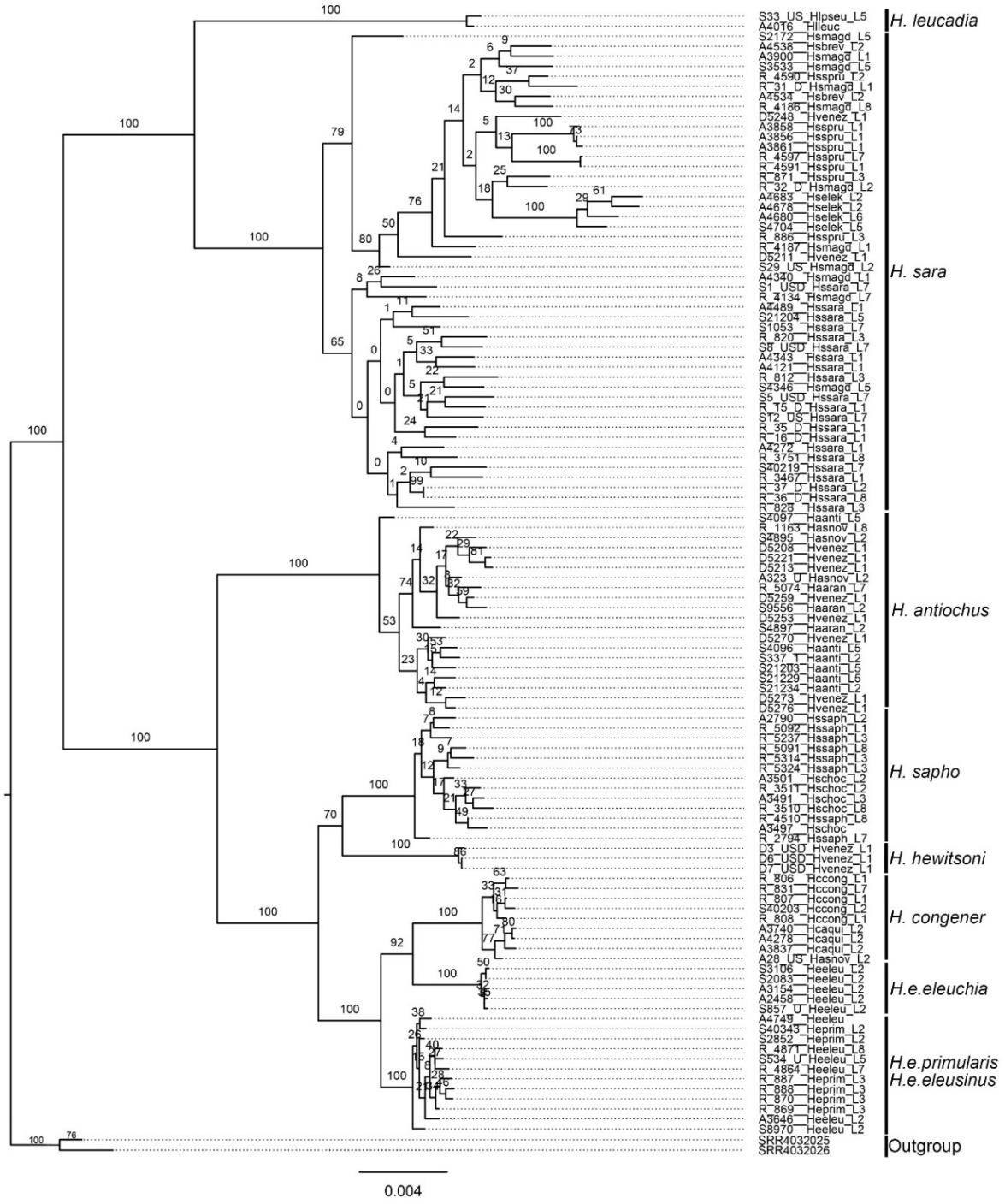

Figure S2.13 Maximum likelihood phylogeny of chromosome 12. Bootstrap support values are indicated at branches.

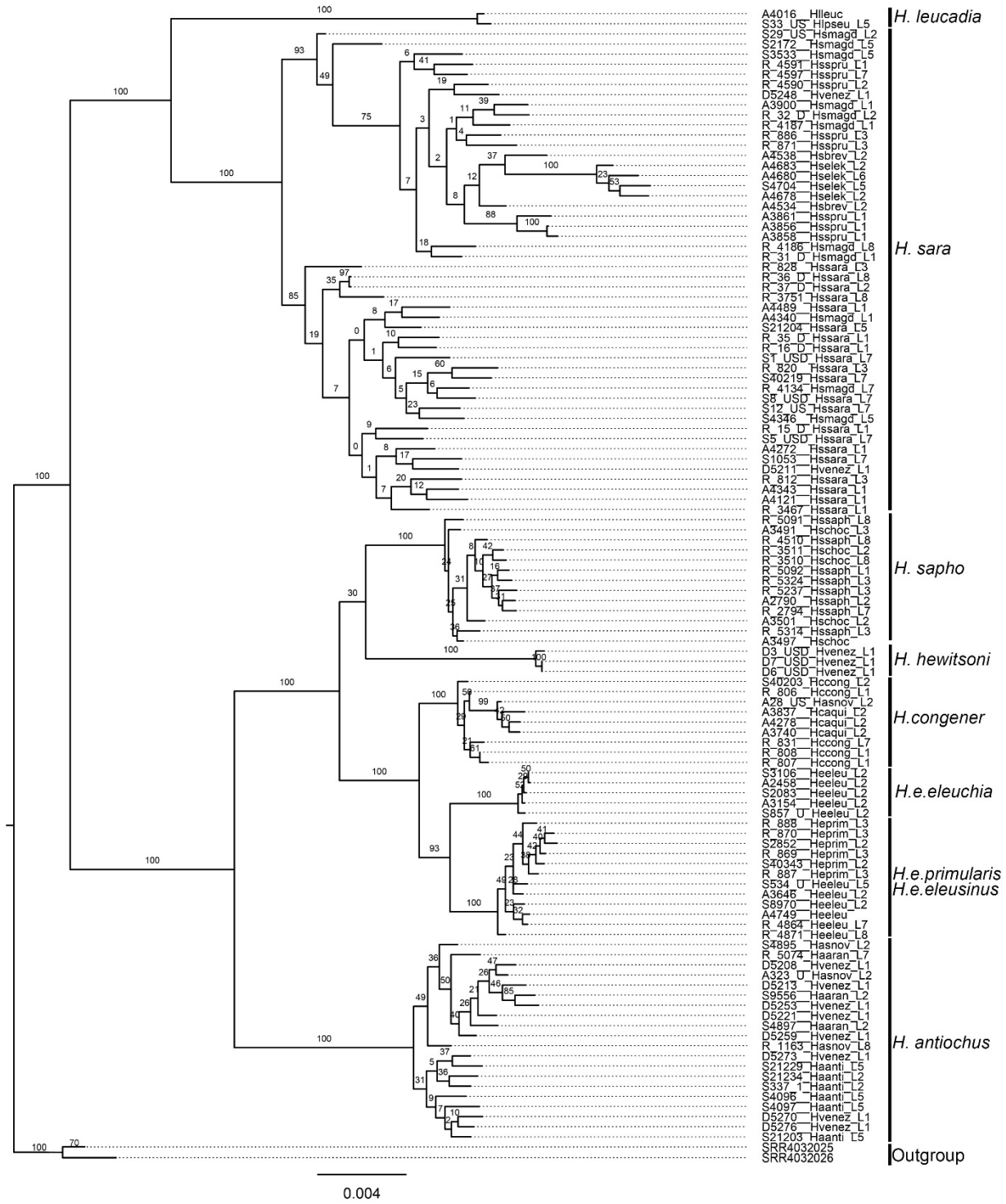

Figure S2.14 Maximum likelihood phylogeny of chromosome 13. Bootstrap support values are indicated at branches.

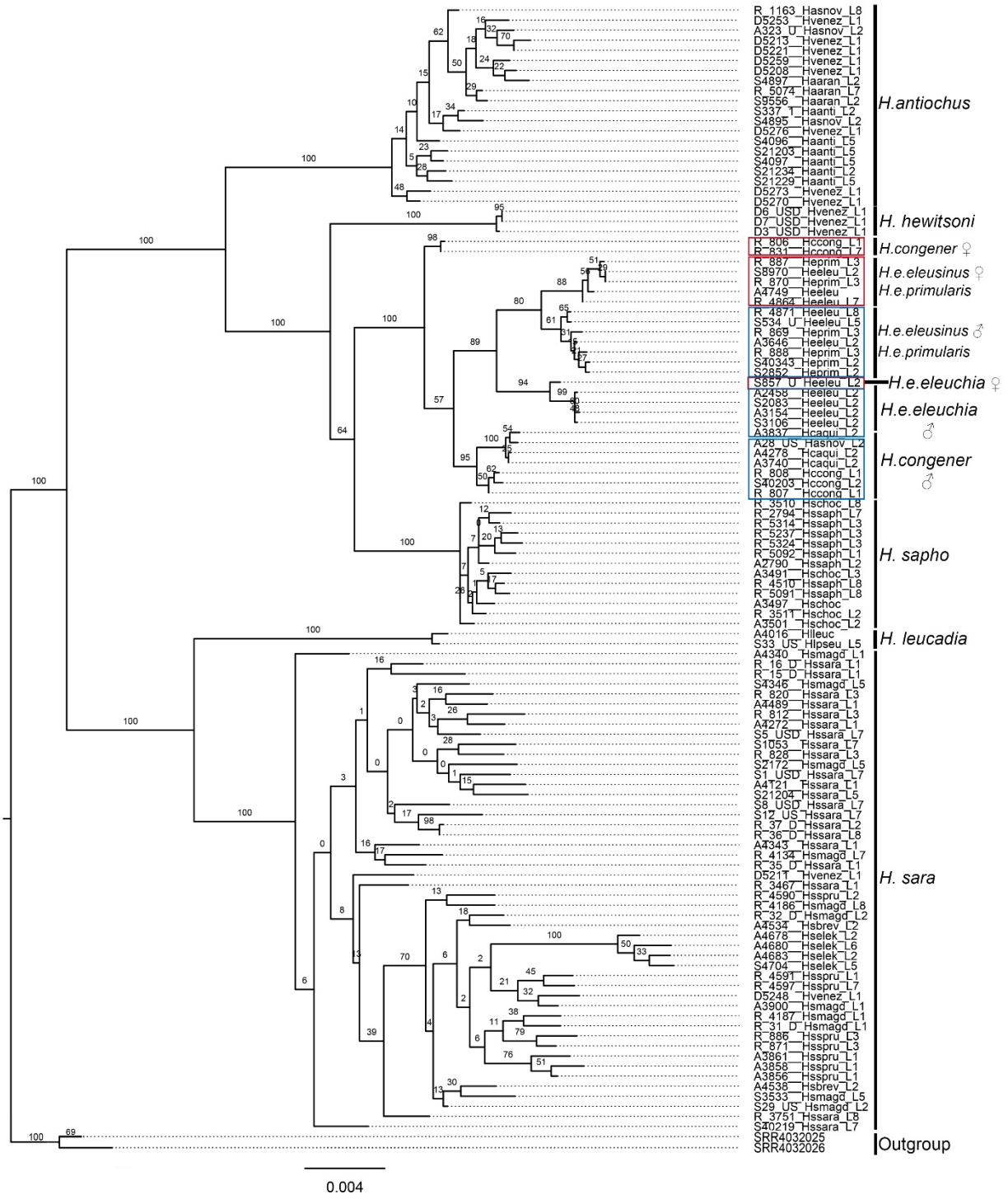

Figure S2.15 Maximum likelihood phylogeny of chromosome 14. Bootstrap support values are indicated at branches. Blue squares group males and while red squares group females.

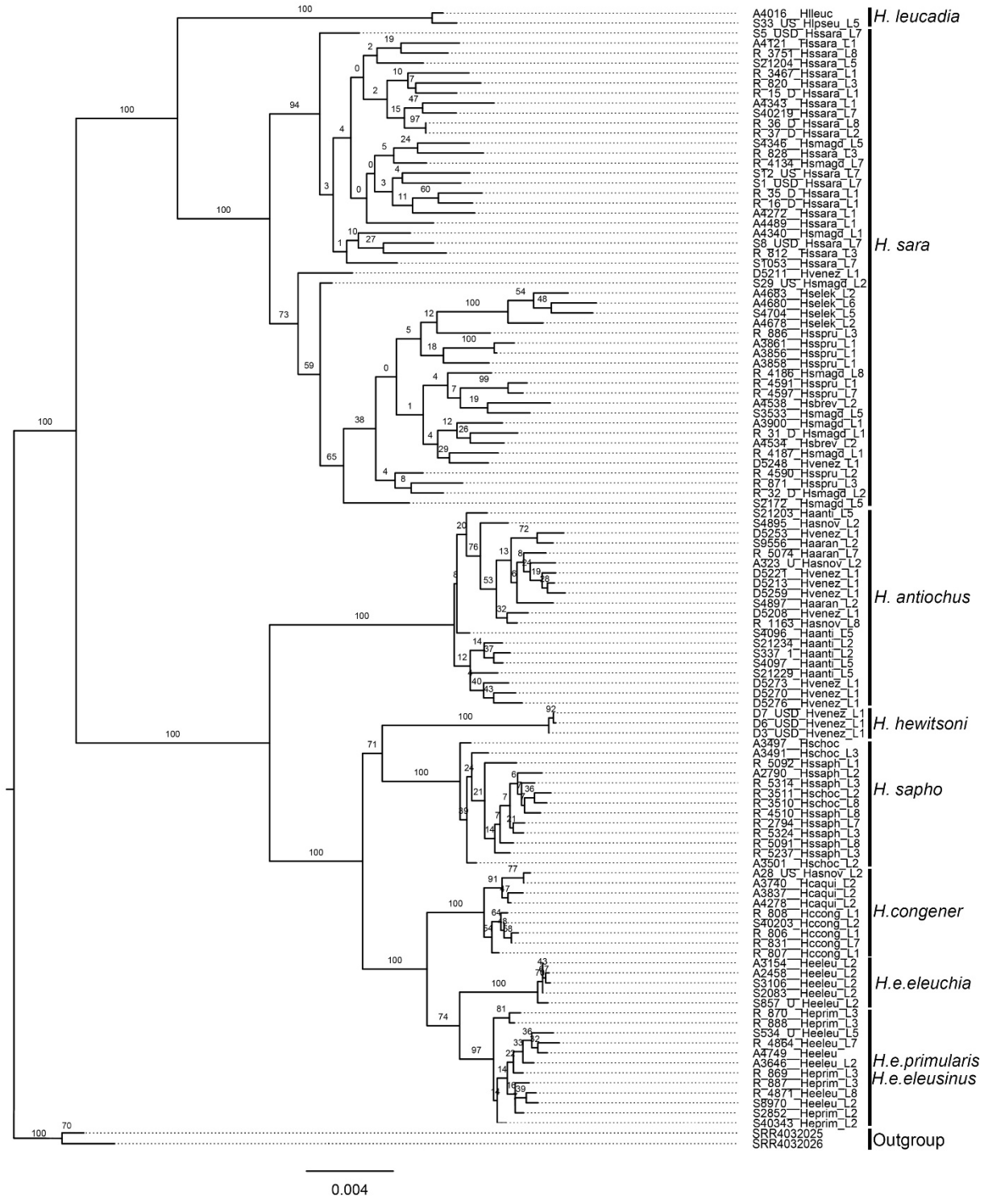

Figure S2.16 Maximum likelihood phylogeny of chromosome 15. Bootstrap support values are indicated at branches.

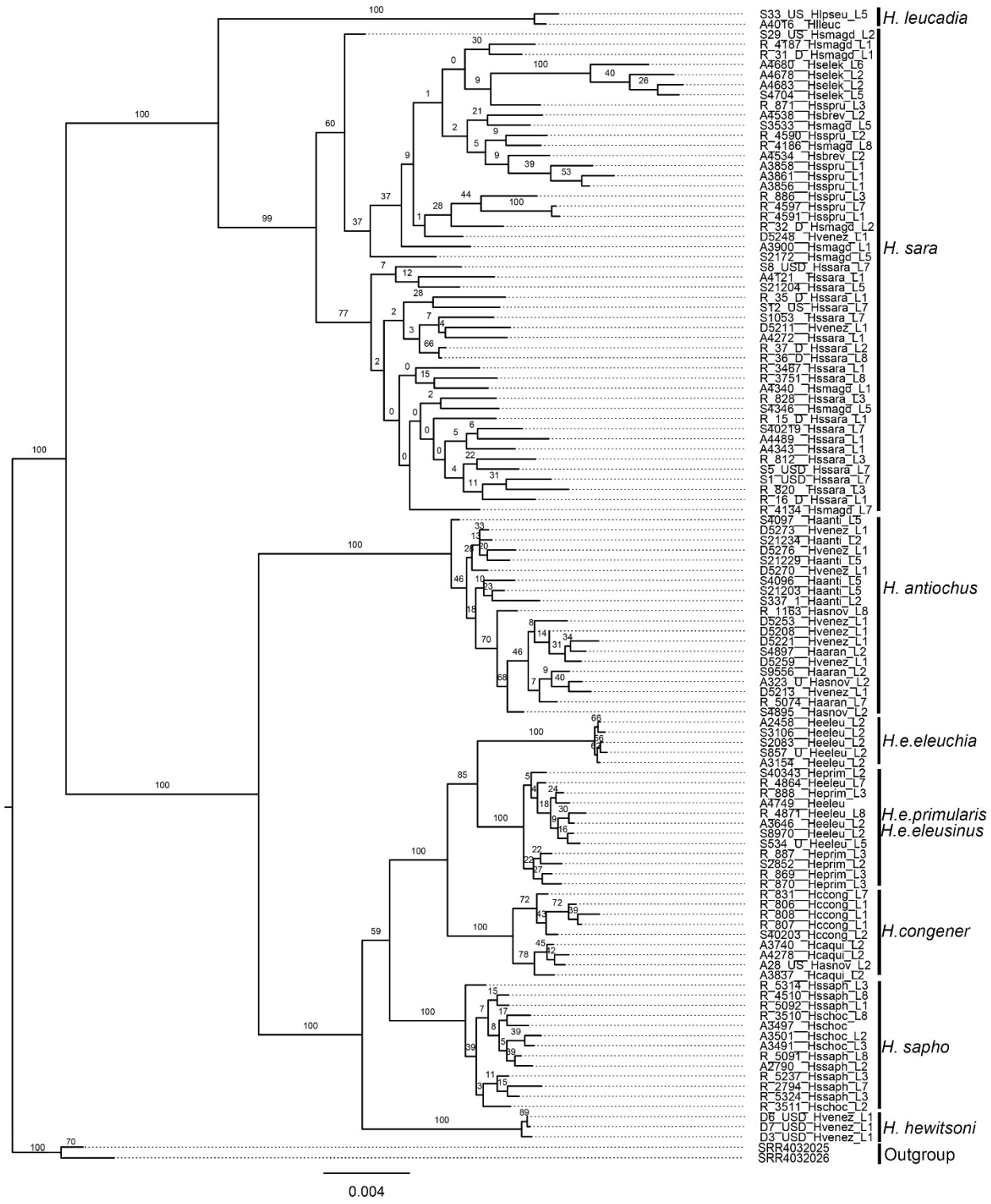

Figure S2.17 Maximum likelihood phylogeny of chromosome 16. Bootstrap support values are indicated at branches.

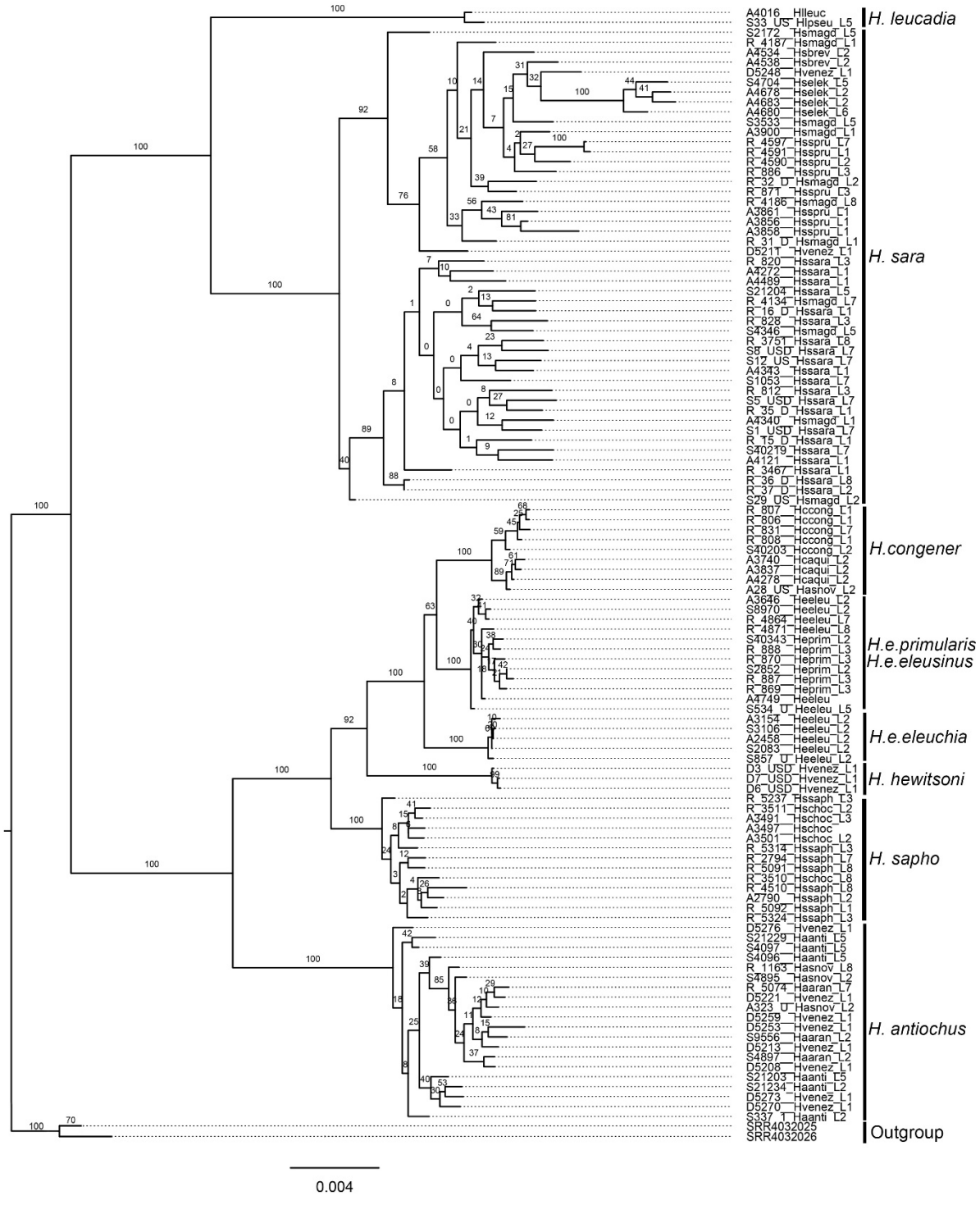

Figure S2.18 Maximum likelihood phylogeny of chromosome 17. Bootstrap support values are indicated at branches.

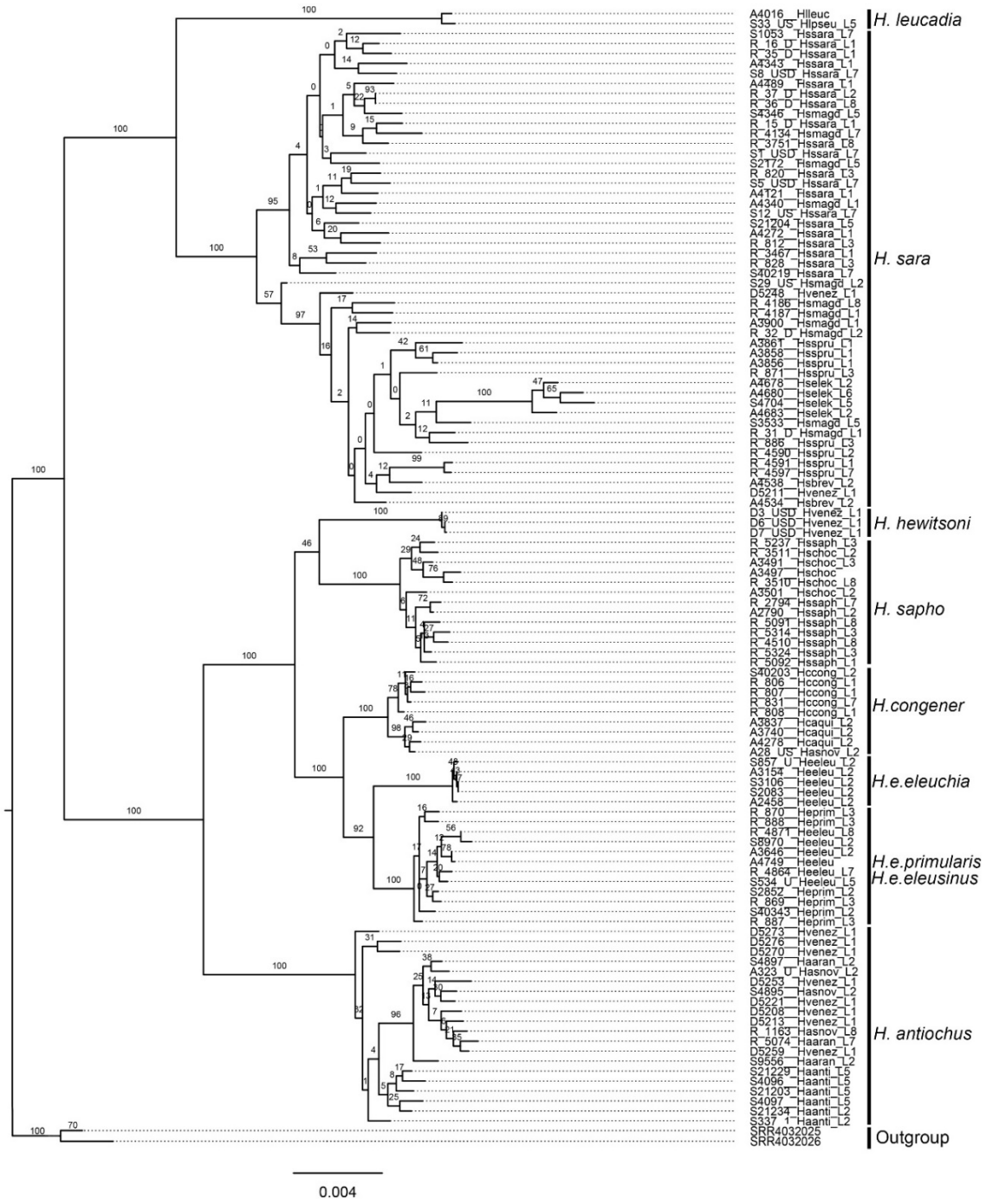

Figure S2.19 Maximum likelihood phylogeny of chromosome 18. Bootstrap support values are indicated at branches, and the scale bar represents the percentage of substitutions per site.

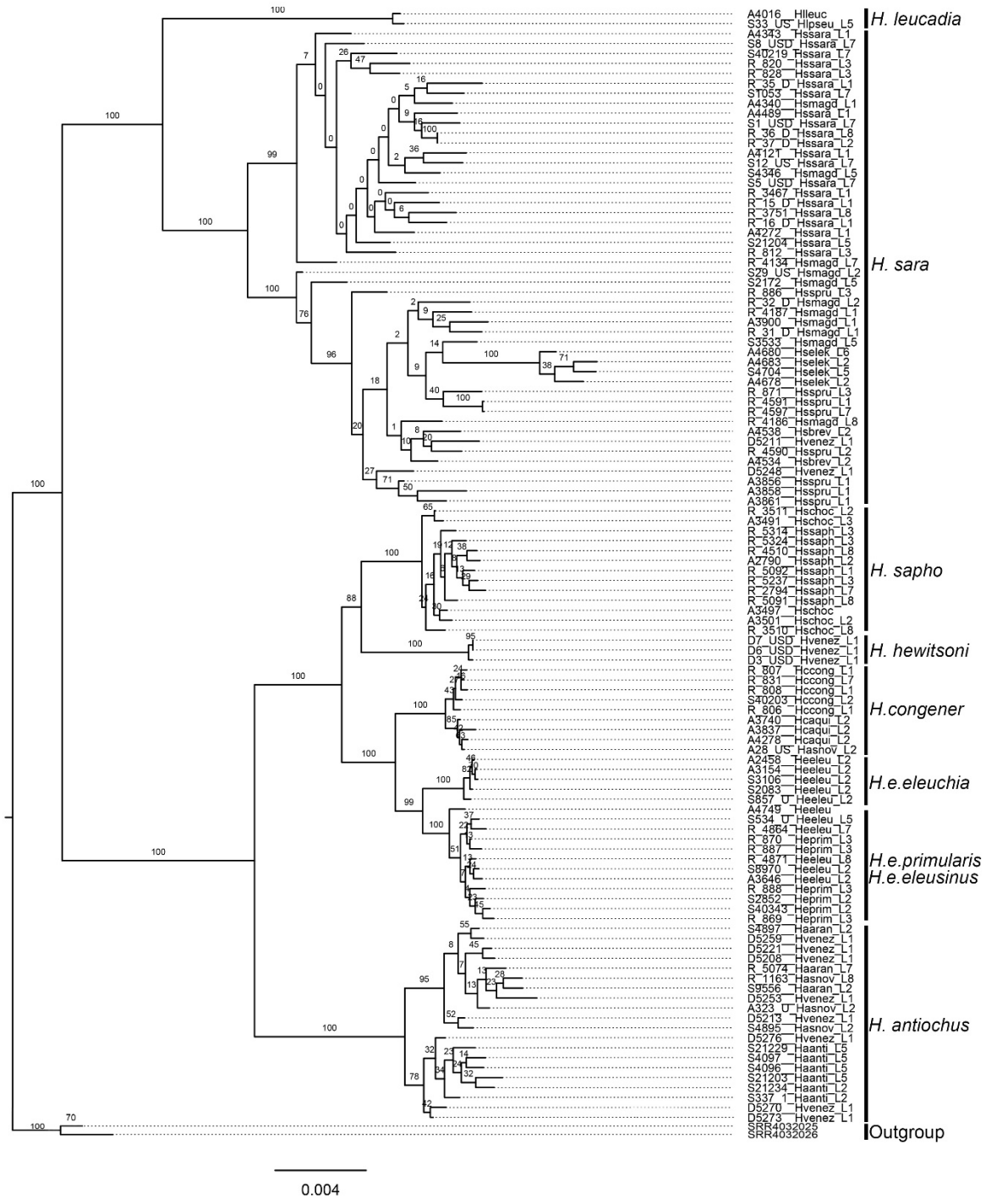

Figure S2.20 Maximum likelihood phylogeny of chromosome 19. Bootstrap support values are indicated at branches.

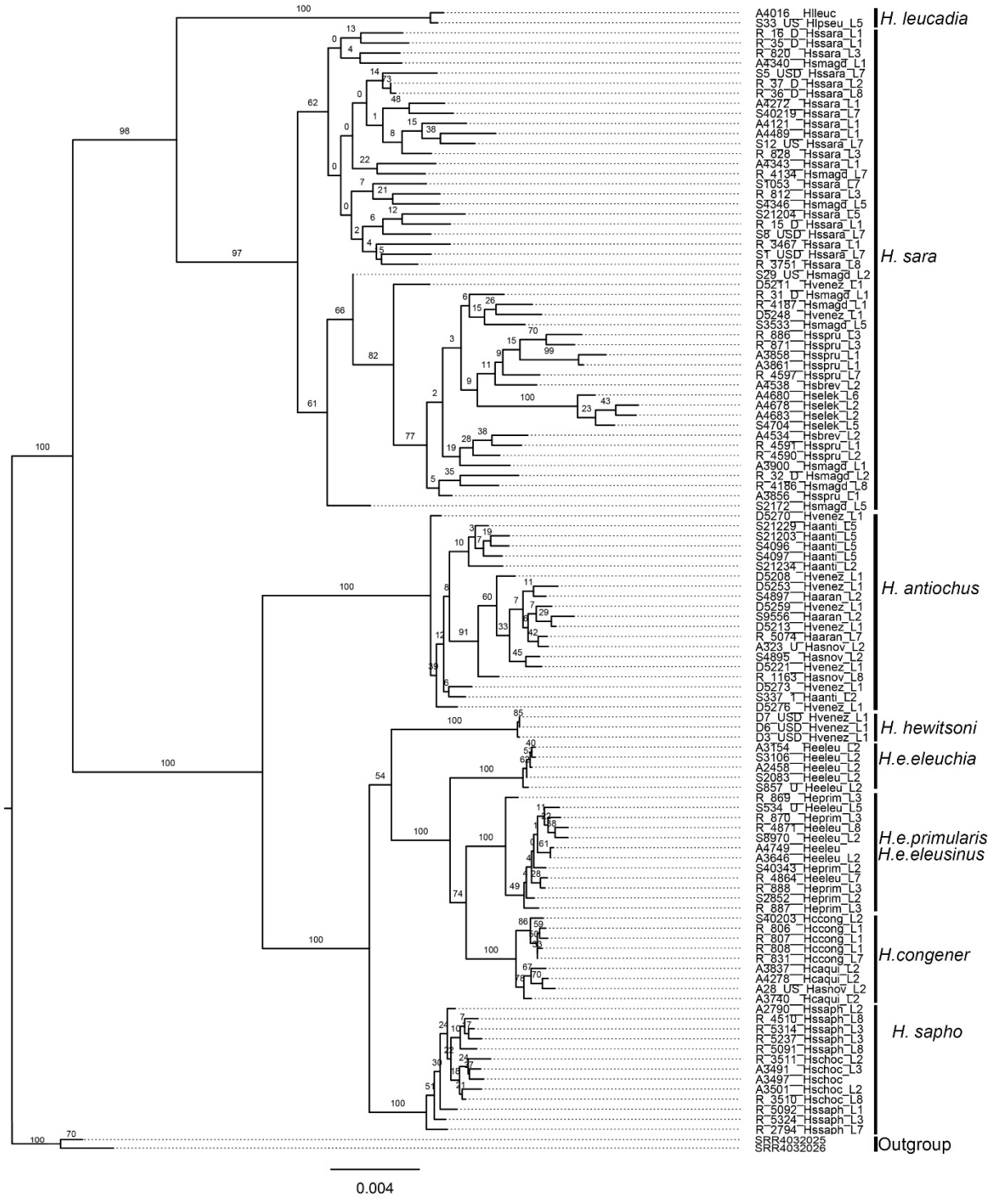

Figure S2.21 Maximum likelihood phylogeny of chromosome 20. Bootstrap support values are indicated at branches.

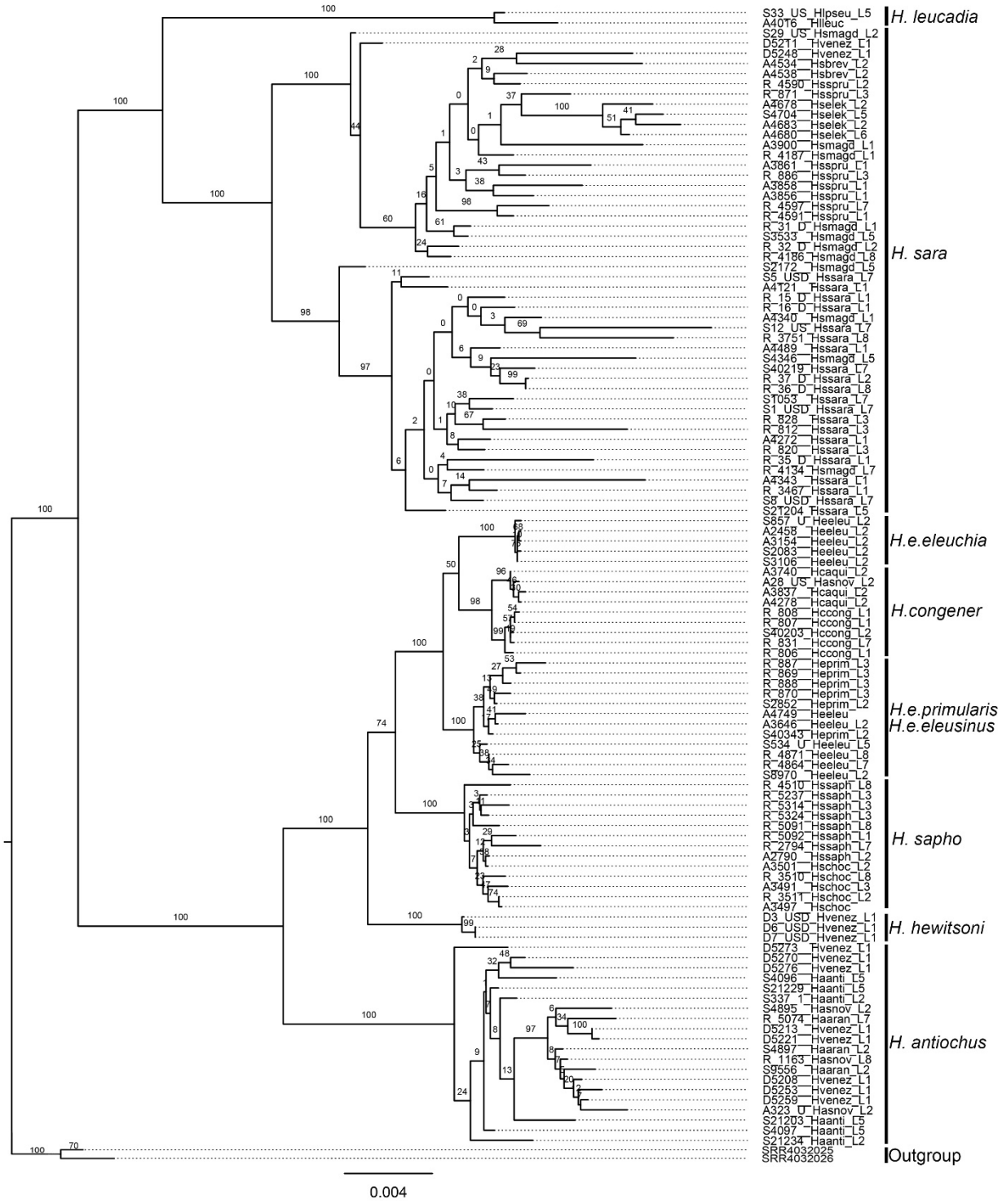

Figure S2.22 Maximum likelihood phylogeny of chromosome 21. Bootstrap support values are indicated at branches.

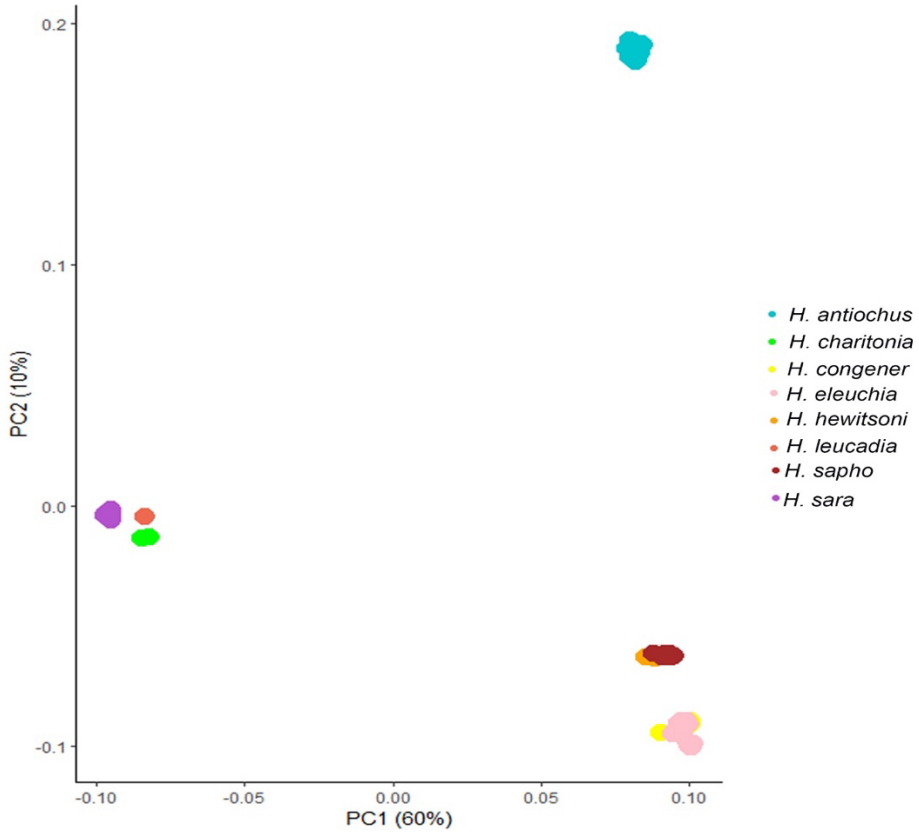

Figure S2.23 Principal Component Analysis (PCA). The analysis was performed using 3.685.916 SNPs.

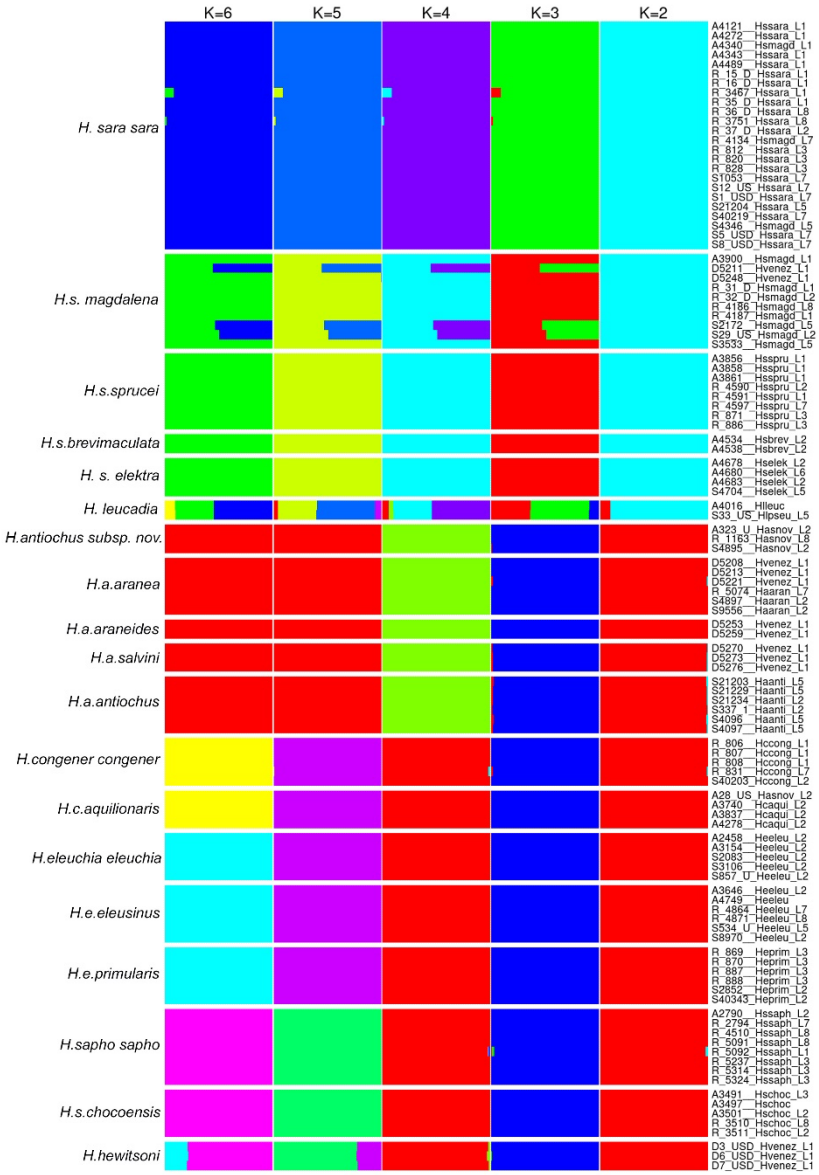

Figure S2.24 Admixture plot. Colored bars indicate genotype cluster proportions for each sample inferred by ADMITXTURE with from K=2 to K = 6.

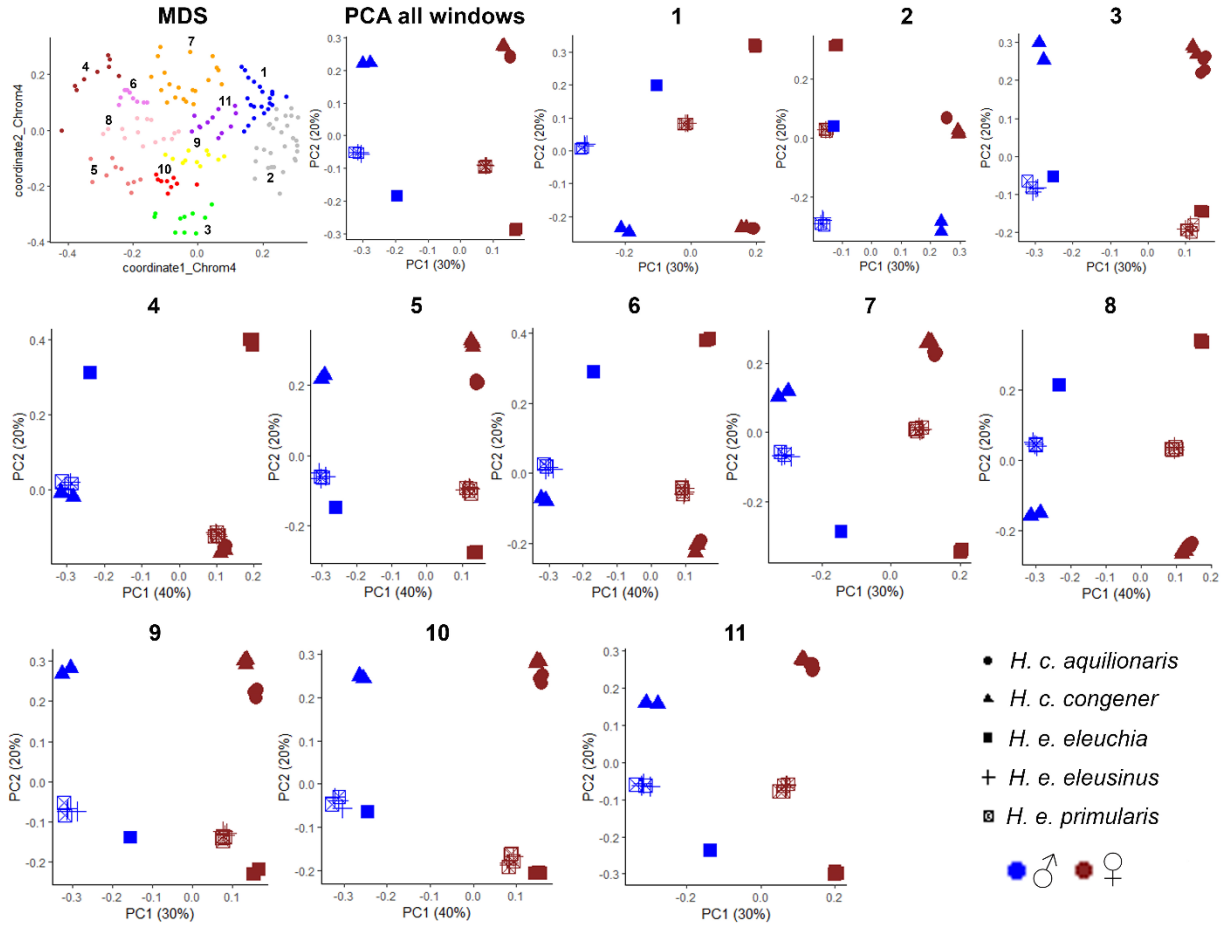

Figure S2.25. Local PCA in non-overlapping windows of 100 SNPs along Chr4 in *H. eleuchia* and *H. congener*. Each group of windows in the MDS analysis (top left) is numbered and coloured coded. Then, PCAs are shown for all windows and for each group of windows derived from the MDS analysis (number indicated on top). In the PCAs each species is represented by a symbol, females are coloured in red, and males are coloured in blue.

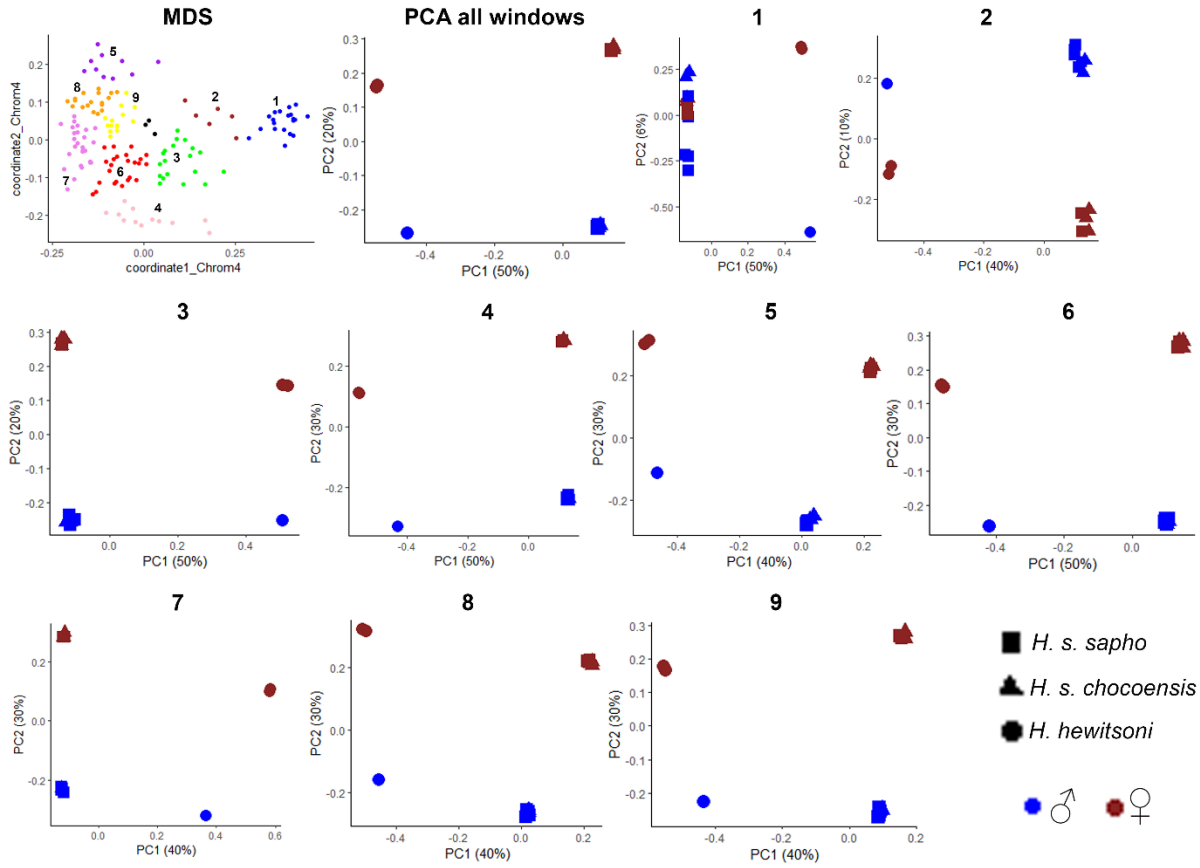

Figure S2.26 Local PCA in non-overlapping windows of 100 SNPs along Chr4 in *H. sapho* and *H. hewitsoni*. Each group of windows in the MDS analysis (top left) is numbered and coloured coded. Then, PCAs are shown for all windows and for each group of windows derived from the MDS analysis (number indicated on top). In the PCAs each species is represented by a symbol, females are coloured in red, and males are coloured in blue.

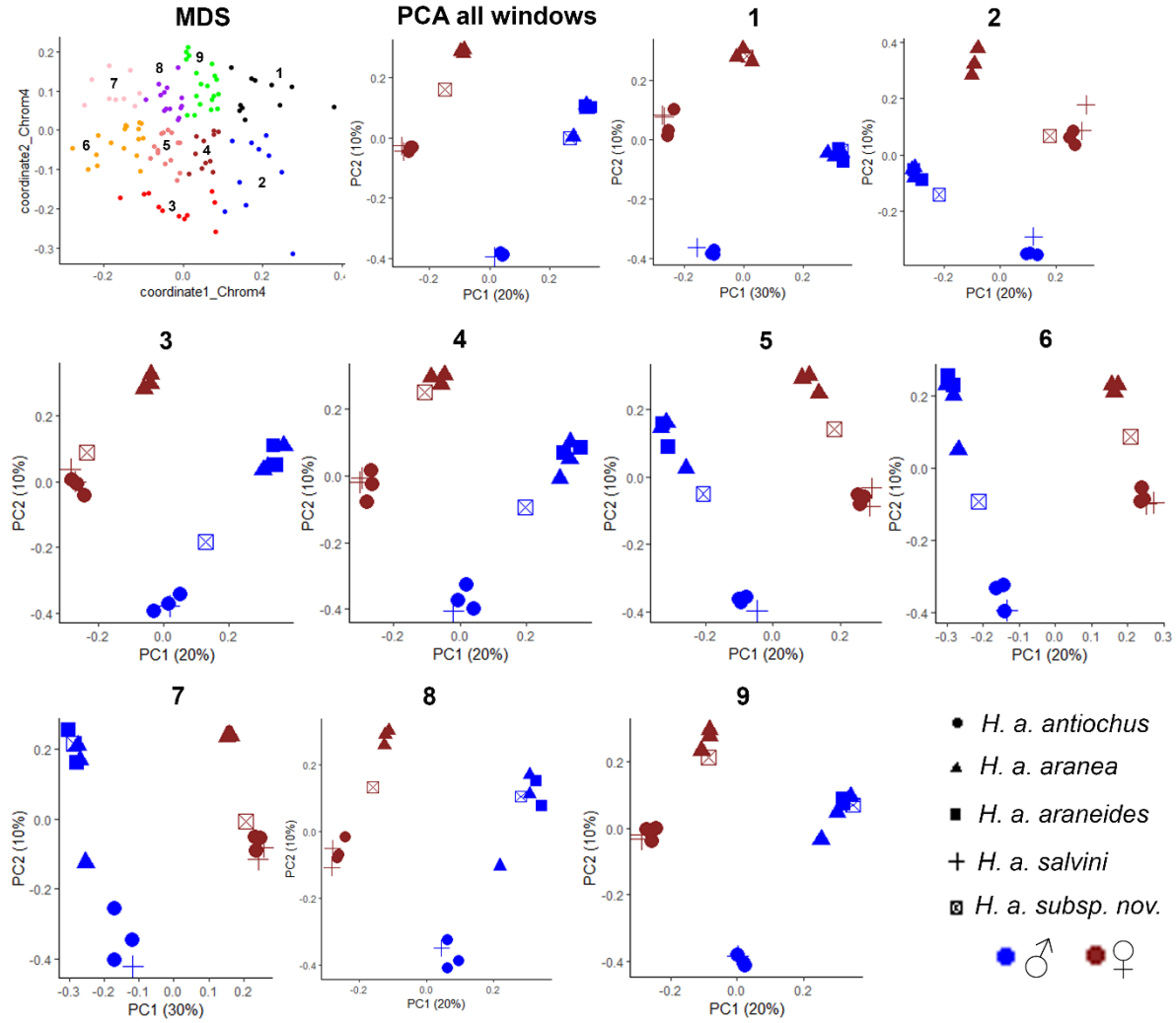

Figure S2.27 Local PCA in non-overlapping windows of 100 SNPs along Chr4 in *H. antiochus*. Each group of windows in the MDS analysis (top left) is numbered and coloured coded. Then, PCAs are shown for all windows and for each group of windows derived from the MDS analysis (number indicated on top). In the PCAs each species is represented by a symbol, females are coloured in red, and males are coloured in blue.

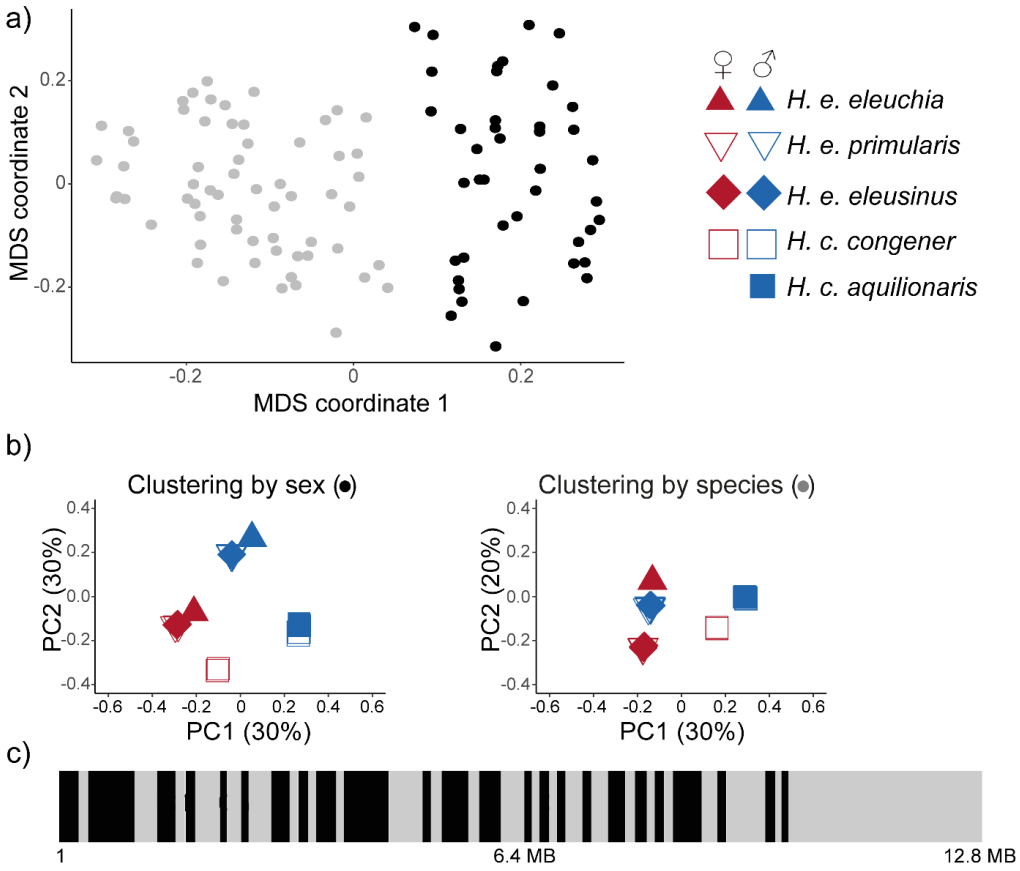

Figure S2.28 Local PCA in non-overlapping windows of 100 SNPs along Chr14 in *H. eleuchia* and *H. congener*. (a) two-dimensional MDS plot of the distribution of 100 SNP windows along Chr14 for the *H. congener/H. eleuchia* clade. Each point represents one window. The left panel shows the relationships between windows, and the middle and right show the two MDS coordinates against position in Chr14. Black: windows that cluster by sex in the local PCA. Grey: windows that cluster by species in the local PCA. (b) PCA plots for the black (left) and grey windows (right) for the *H. congener/H. eleuchia* clade. Each species is represented by a symbol, and each individual is represented with a point. The female symbol for *H. congener aquilionaris* is missing because there are no female specimens of that subspecies available.

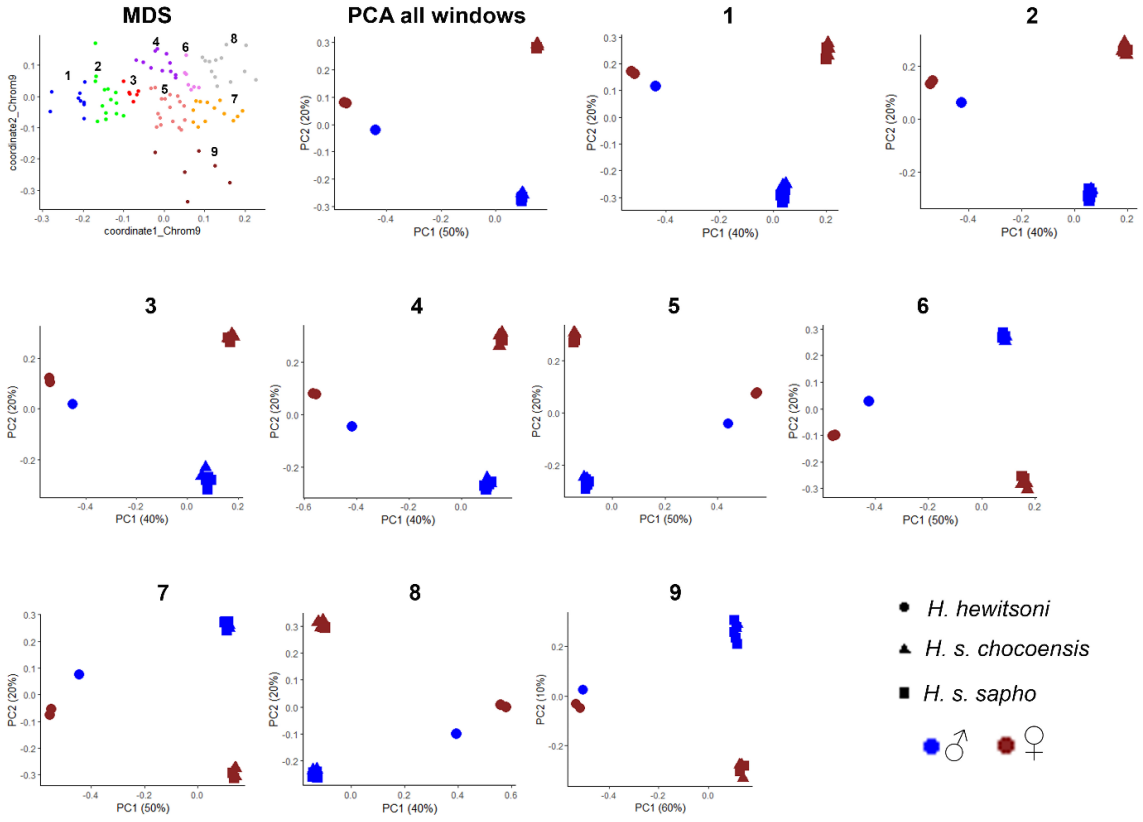

Figure S2.29 Local PCA in non-overlapping windows of 100 SNPs along Chr9 in *H. sapho* and *H. hewitsoni*. Each group of windows in the MDS analysis (top left) is numbered and coloured coded. Then, PCAs are shown for all windows and for each group of windows derived from the MDS analysis (number indicated on top). In the PCAs each species is represented by a symbol, females are coloured in red, and males are coloured in blue.

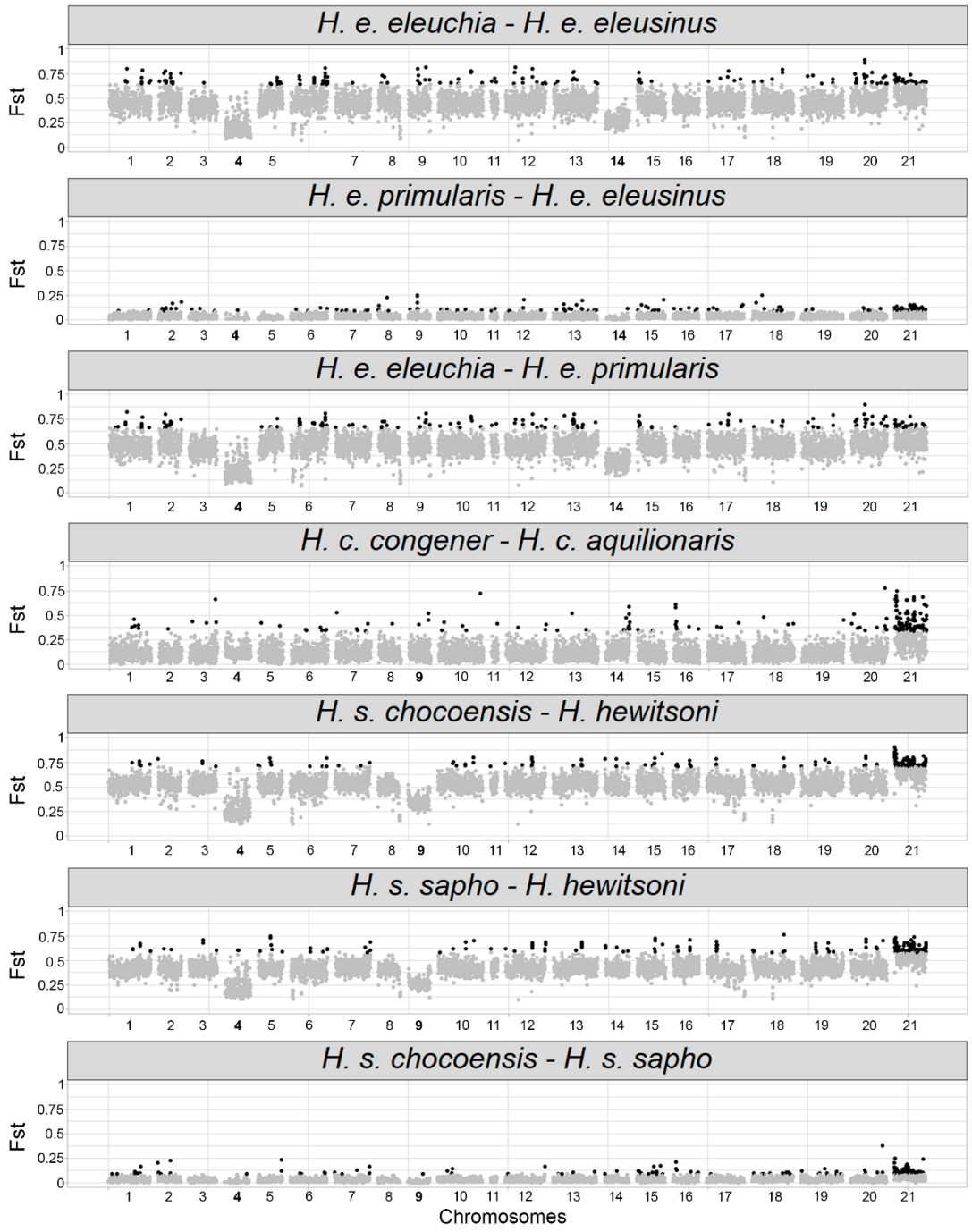

Figure S2.30 Genome-wide divergence (FST) between pairs of subspecies of *H. eleuchia*, *H. congener* and *H. sapho*. Each point represents a 50Kb window. The significance threshold is set at the top 5% of the F_{ST} values distribution tail, and black windows are those that passed this threshold.

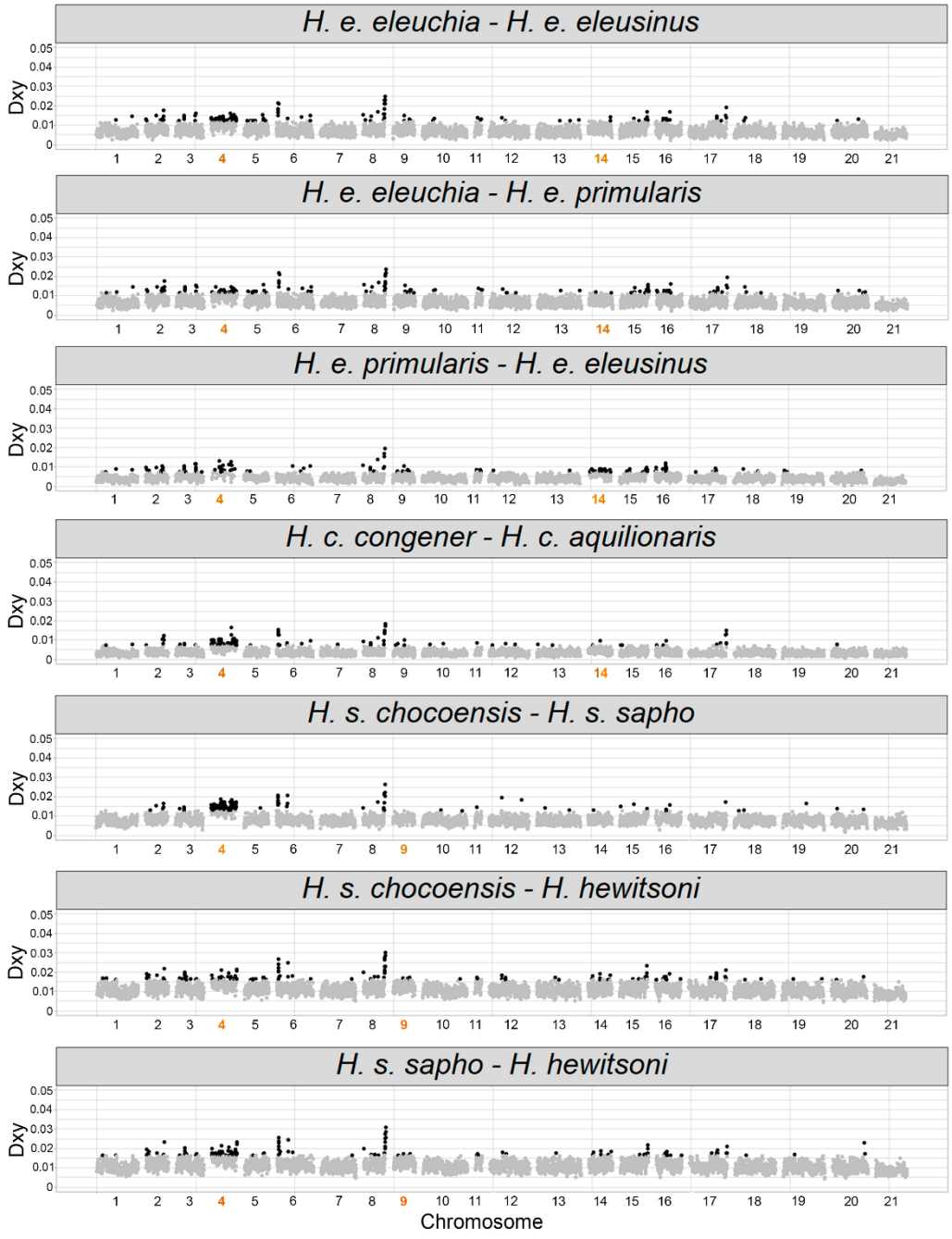

Figure S2.31 Genome-wide divergence (Dxy) between pairs of all subspecies in the *sara/sapho* clade. Each point represents a 50Kb window. The significance threshold is set at the top 5% of the F_{ST} values distribution tail, and black windows are those that passed this threshold. Analyses for other pairs of subspecies are shown in Figures S32 and S33.

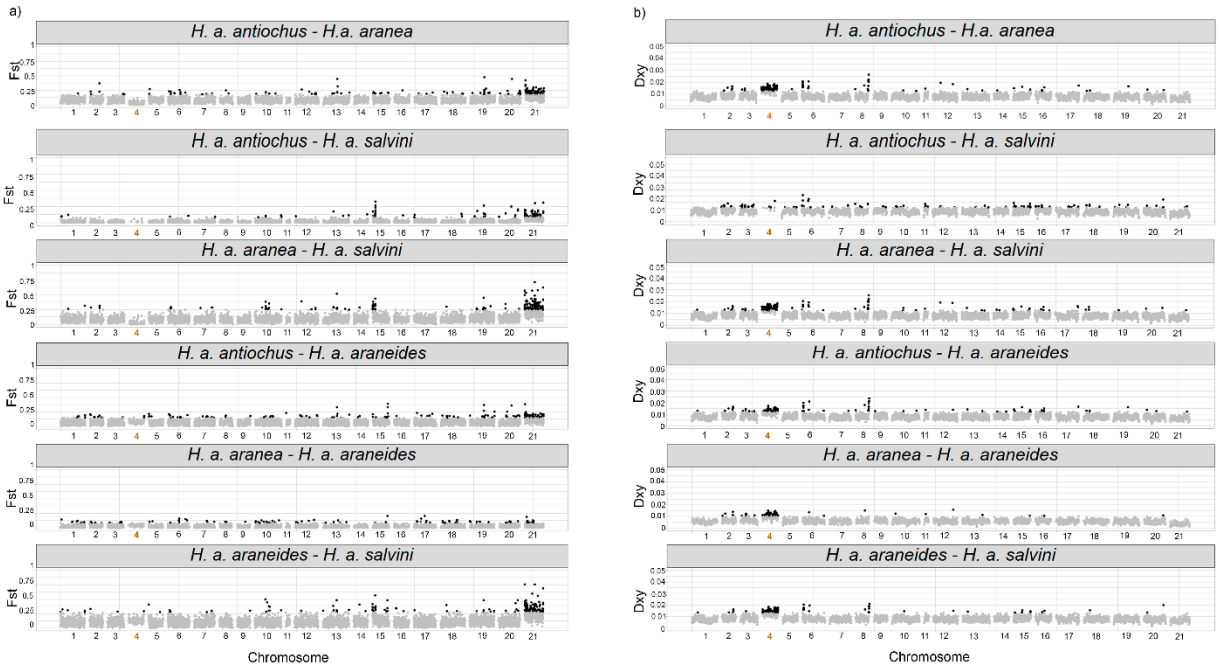

Figure S2.32 Genome-wide divergence between pairs of subspecies of *H. antiochus*. (a) F_{ST} and (b) Dxy. Each point represents a 50Kb window. The significance threshold is set at the top 5% of the F_{ST} values distribution tail, and black windows are those that passed this threshold.

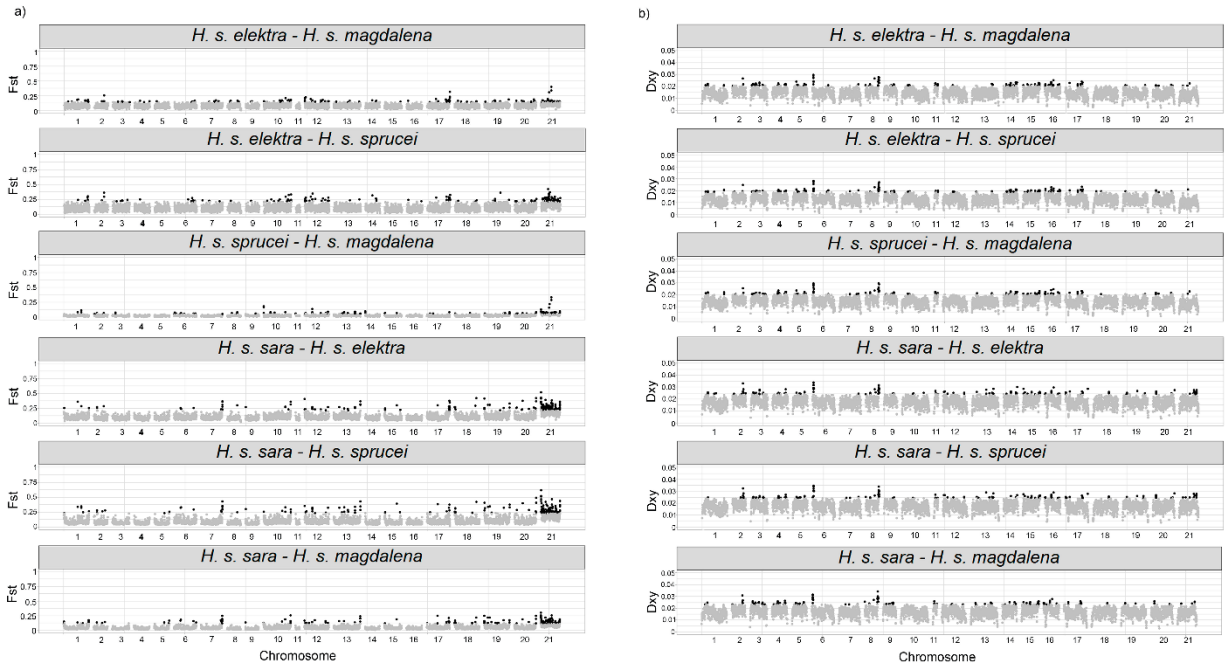

Figure S2.33 Genome-wide divergence between pairs of subspecies of *H. sara*. (a) F_{ST} and (b) Dxy. Each point represents a 50Kb window. The significance threshold is set at the top 5% of the F_{ST} values distribution tail, and black windows are those that passed this threshold.

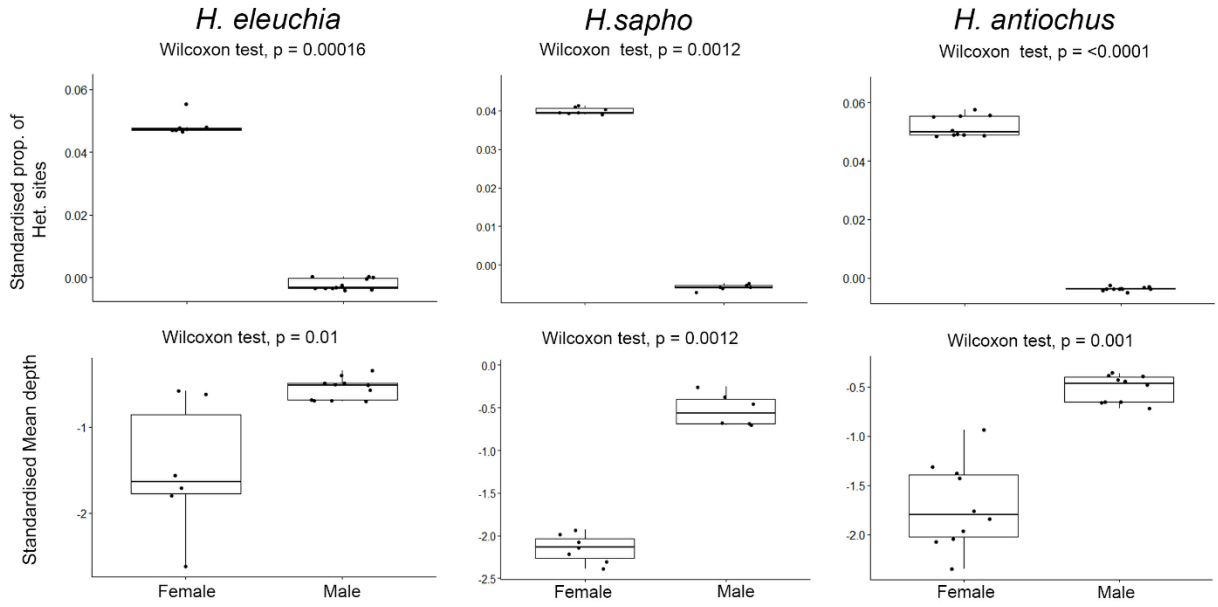

Figure S2.34 Standardised proportion of heterozygous sites and mean depth between sexes on chromosome 4. Each panel corresponds to a species with the standardized proportion of heterozygous sites shown on top and mean depth on the bottom. Each dot represents one individual. ns= non-significant.

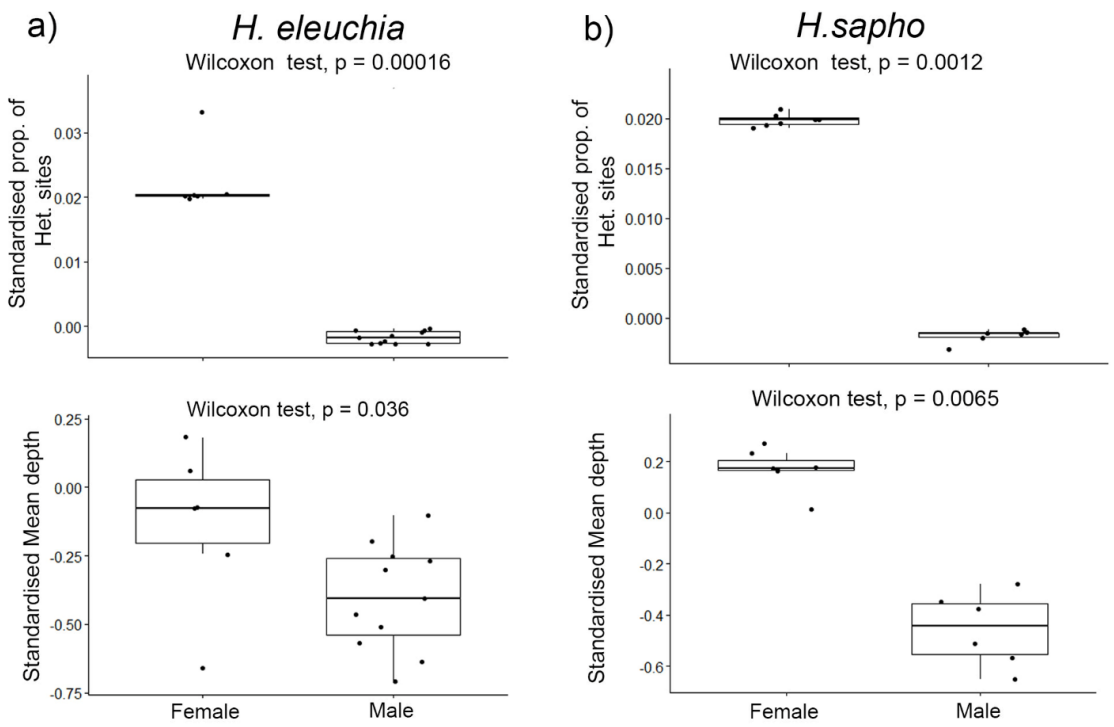

Figure S2.35 Standardised proportion of heterozygous sites and mean depth between sexes on a) 14 y b) 9 chromosomes. Each panel corresponds to a species with the standardized proportion of heterozygous sites shown on top and mean depth on the bottom. Each dot represents one individual. ns= non-significant.

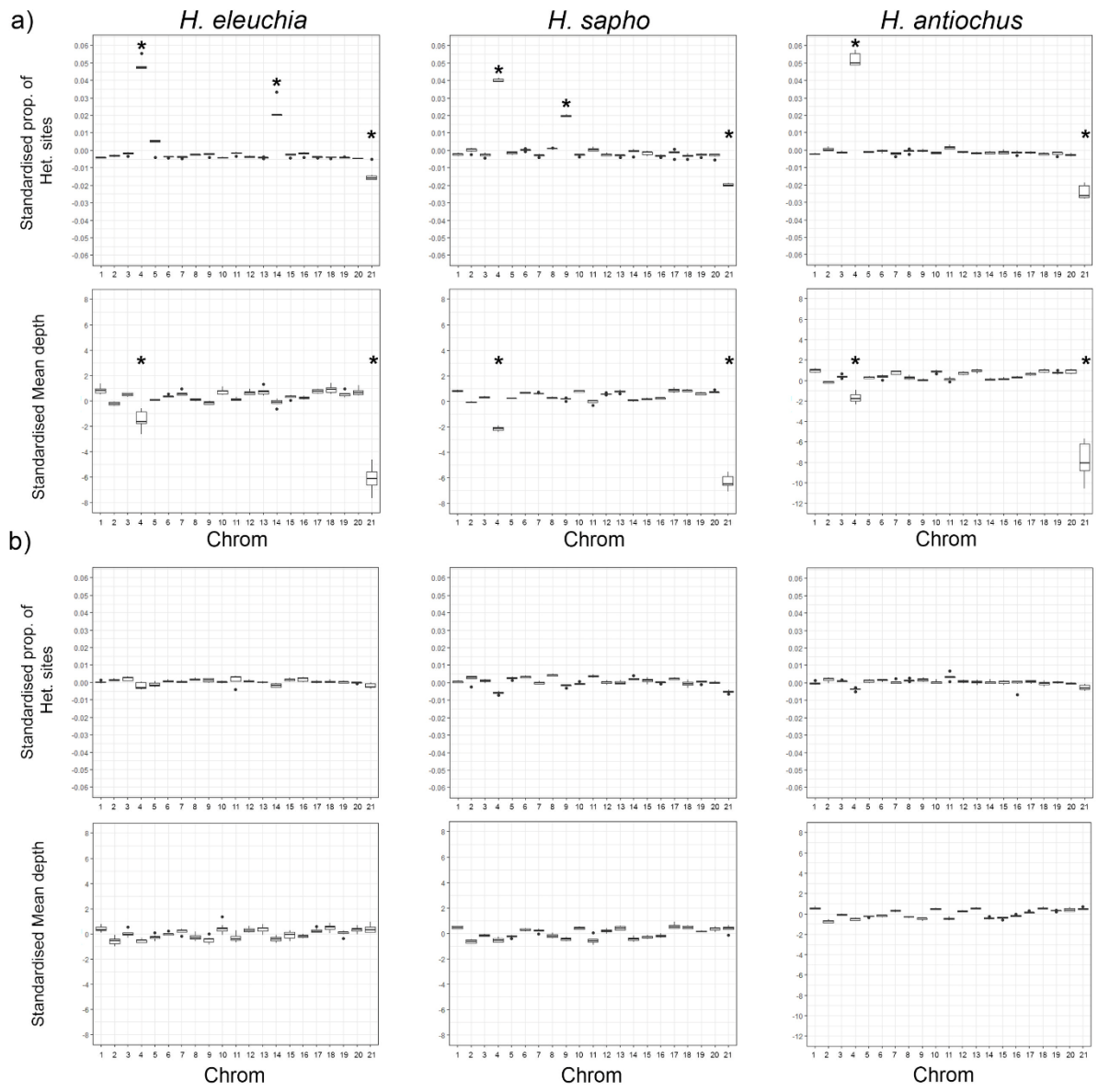

Figure S2.36 Standardised proportion of heterozygous sites and mean depth compared among chromosomes in a) females and b) males of *H. eleuchia, H sapho and H. antiochus* species. Each panel corresponds to a species with the standardized proportion of heterozygous sites shown on top and mean depth on the bottom. Chromosomes with * ($p < 0.01$) are significantly different from all other chromosomes.

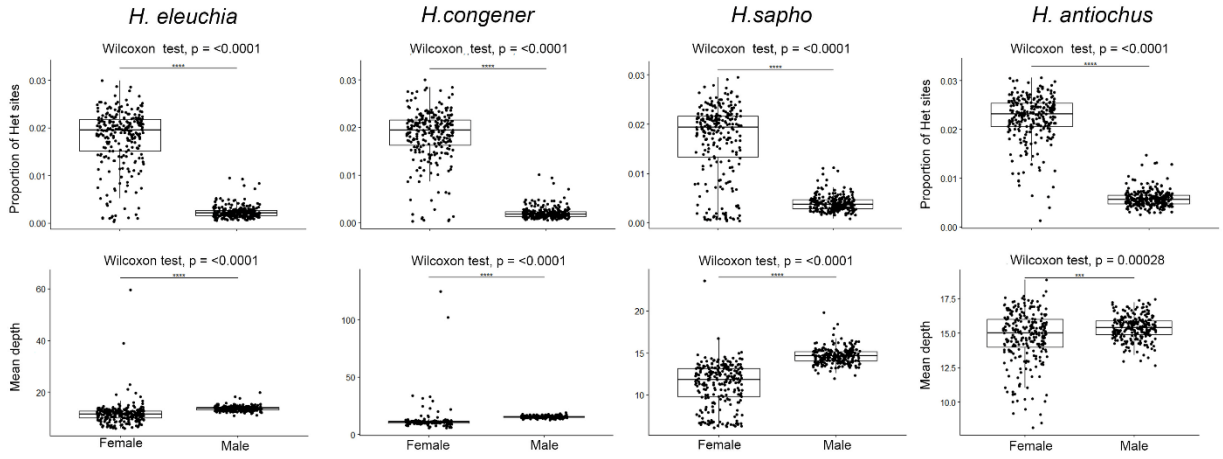

Figure S2.37 Proportion of heterozygous sites and mean depth between sexes in sliding windows along chromosome 4. Each panel corresponds to a species with the proportion of heterozygous sites shown on top and mean depth on the bottom. Each dot represents the average of these values across all individuals per window. ns = non-significant.

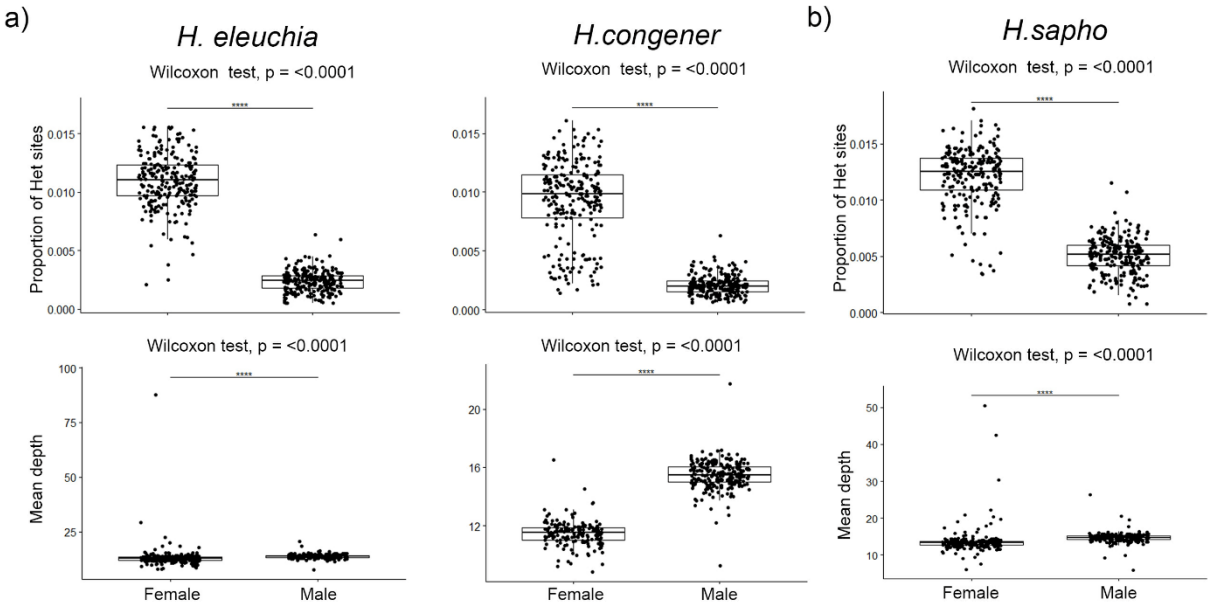

Figure S2.38 Proportion of heterozygous sites and mean depth between sexes in sliding windows along chromosome a) 14 and b) 9. Each panel corresponds to a species with the proportion of heterozygous sites shown on top and mean depth on the bottom. Each dot represents the average of these values across all individuals per window. ns = non-significant.

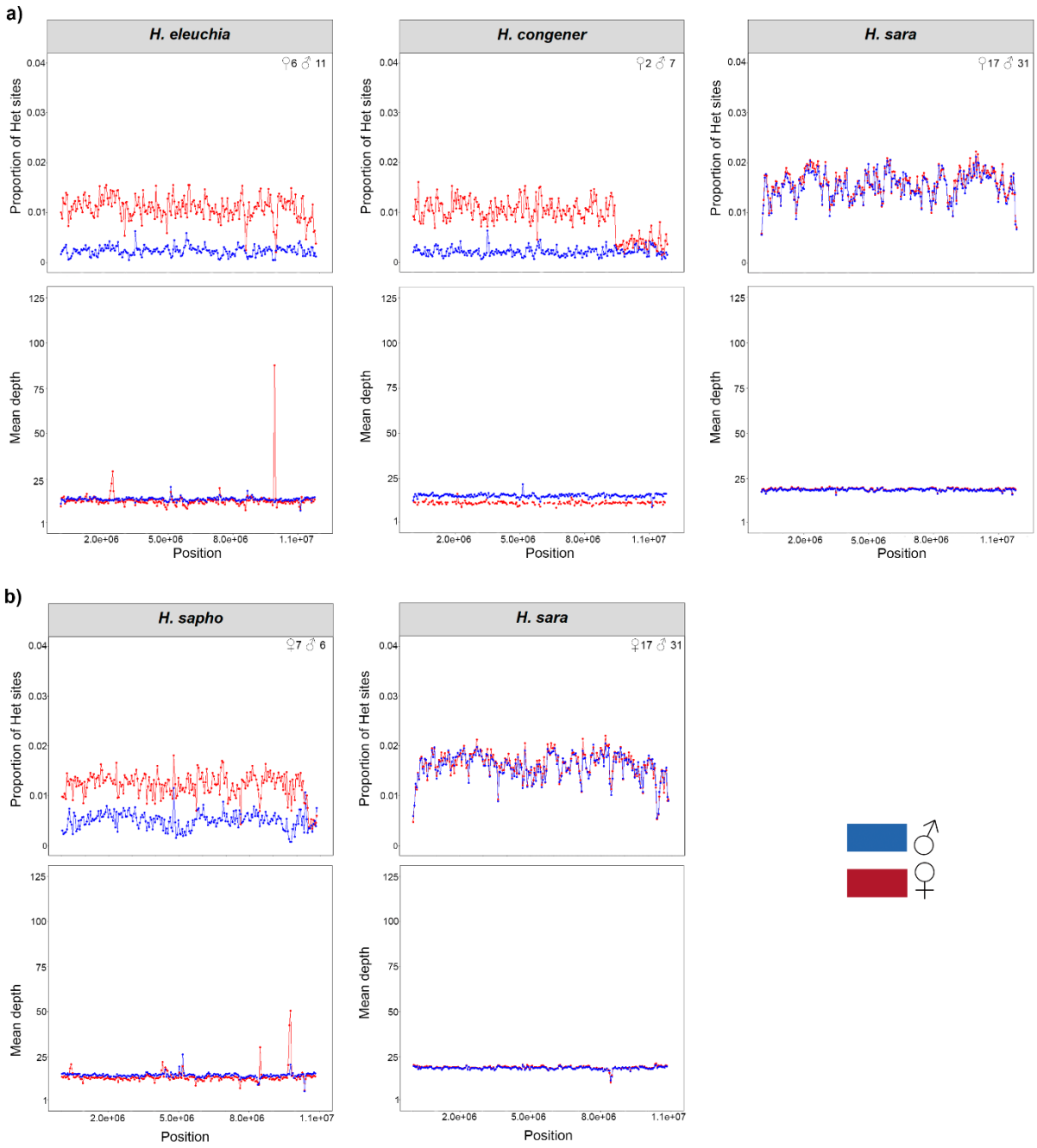

Figure S2.39 Patterns of heterozygosity and depth across chromosome a) 14 and b) 9. Proportion of heterozygosity sites and mean depth in 50Kb sliding windows in each species. Each line corresponds to one individual where males are shown in blue and females in red, and their *n* is shown in the top right corner.

CHAPTER 3

Chemical defence variation in *Heliconius* butterflies: testing the role of mimicry rings and ecoregions

Table S3.1 Sample information and concentration of each CNglc per sample (in ug/mg). Gyn.: Gynocardin, Dih.: Dihydrogynocardin, Tet.: Tetraphyllin B, Epi.: Epivolkenin, Dei.: Deidaclin, Lin.: Linamarin, Lot.: Lotaustralin, Epil.: Epilotaustralin. Total: concentration of all CNglcs combined.

ID	Ring	Species	Subspecies	Locality	Gyn.	Dih.	Tet.	Epi.	Dei.	Lin.	Lot.	Epil.	Total
4128	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas	0.00	0.00	0.00	8.02	0.00	4.77	2.99	0.44	16.23
4129	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas	0.00	0.00	0.00	24.29	0.00	0.25	0.11	0.00	24.66
4130	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas	0.00	0.00	0.00	32.40	0.00	0.09	0.03	0.00	32.52
5029	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas	0.00	0.00	0.00	0.00	0.00	15.42	6.69	0.00	22.11
5030	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas	0.00	0.00	0.00	90.61	0.04	0.68	0.07	0.00	91.41
5341	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	11.31	5.35	3.06	19.72
5363	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	1.34	31.65	19.73	16.85	69.58
5364	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas-Pedrera	0.00	0.00	0.00	60.04	0.00	1.78	0.85	0.03	62.69
5383	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	35.77	19.49	10.91	66.17
5384	No mimicry ring	*H. antiochus*	*antiochus*	Amazonas-Pedrera	0.00	0.00	0.00	49.85	0.00	0.87	1.02	0.01	51.75
5021	Dennis-ray	*H. aoede*	*bartleti*	Amazonas	0.00	0.00	0.00	65.68	0.00	0.00	0.00	0.00	65.68
5022	Dennis-ray	*H. aoede*	*bartleti*	Amazonas	0.00	0.00	0.00	51.52	0.00	0.00	0.00	0.00	51.52
4208	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	10.26	0.00	0.00	0.00	0.00	1.89	1.61	0.07	13.84

4209	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	44.75	0.00	1.87	0.00	0.00	9.17	5.08	0.97	61.84
4211	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	5.54	0.00	0.00	0.00	0.00	4.00	2.05	0.40	11.98
4212	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	19.16	0.00	0.00	0.00	0.00	4.54	1.80	0.17	25.66
4213	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	10.33	0.00	0.00	0.00	0.00	2.75	1.24	0.16	14.49
5588	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	0.00	0.00	0.00	0.00	0.00	4.30	1.83	0.95	7.08
5589	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	2.07	0.00	0.00	0.00	0.00	0.98	0.94	0.00	3.99
5590	No mimicry ring	*H. charithonia*	*bassleri*	Boyaca	0.00	0.00	0.00	0.00	0.00	3.01	1.13	0.96	5.11
4645	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca valley	0.00	0.00	0.00	2.06	0.36	8.63	3.32	0.46	14.82
4646	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	1.07	0.00	7.28	3.63	2.02	14.00
4664	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	3.82	0.00	10.07	4.67	1.71	20.26
4665	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	14.16	0.00	0.00	6.15	1.96	0.17	22.44
4666	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	33.69	0.00	3.23	0.40	0.00	37.31
5674	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	13.27	5.19	0.00	18.46
5675	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	6.94	2.76	0.94	10.64
5676	No mimicry ring	*H. clysonimus*	*clysonimus*	Cauca Valley	0.00	0.00	0.00	42.33	0.00	0.00	0.00	0.00	42.33
5007	White_yellow	*H. cydno*	*cydnides*	Caldas	0.00	0.00	0.72	0.00	1.84	14.35	7.70	0.00	24.61
5008	White_yellow	*H. cydno*	*cydnides*	Caldas	0.00	0.00	0.00	0.00	0.00	5.26	1.11	0.00	6.37
5010	White_yellow	*H. cydno*	*cydnides*	Caldas	0.30	0.31	0.00	0.00	0.00	1.20	0.14	0.00	1.94
5011	White_yellow	*H. cydno*	*cydnides*	Caldas	0.24	0.30	0.00	0.00	0.00	0.84	0.10	0.00	1.49
5485	White_yellow	*H. cydno*	*cydnides*	Cauca Valley	0.00	0.30	0.00	0.00	0.00	14.51	7.87	5.57	28.25
5486	White_yellow	*H. cydno*	*cydnides*	Cauca Valley	0.00	0.00	0.00	0.00	6.00	4.96	2.15	0.58	13.69
5487	White_yellow	*H. cydno*	*cydnides*	Cauca Valley	0.00	1.68	0.00	0.00	0.00	10.91	4.72	4.15	21.46
5488	White_yellow	*H. cydno*	*cydnides*	Cauca Valley	0.00	0.00	0.00	0.00	7.58	12.96	6.50	6.70	33.73
4792	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	4.76	2.03	2.64	9.42
4793	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	2.78	9.57	3.97	0.00	16.32
4794	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	6.26	2.15	0.50	8.92
4795	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	4.72	2.69	0.22	7.63

4796	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.90	6.66	2.12	1.20	10.88
5666	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	9.25	2.29	0.65	12.18
5667	Yellow_hindwing	*H. cydno*	*weymeri*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	16.65	5.80	6.94	29.40
5780	White	*H. cydno*	*zelinde*	Buenaventura	0.00	0.00	0.00	0.00	0.00	27.57	12.45	7.11	47.13
5789	White	*H. cydno*	*zelinde*	Buenaventura	0.00	0.00	0.00	1.95	0.00	21.89	7.99	0.00	31.83
4622	Yellow	*H. doris*	*doris*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	29.98	13.12	2.31	45.41
4663	Yellow	*H. doris*	*doris*	Cauca Valley	0.00	0.00	0.00	0.00	0.25	23.79	0.70	0.00	24.74
4669	Yellow	*H. doris*	*doris*	Cauca Valley	0.00	0.00	0.00	0.00	0.00	24.76	1.12	0.06	25.95
5537	Yellow	*H. doris*	*doris*	Amazonas	0.00	0.00	0.00	0.00	0.00	7.28	4.61	0.24	12.12
5636	Yellow	*H. doris*	*doris*	Boyaca	0.00	13.29	0.00	0.00	0.00	57.60	24.93	0.00	95.82
5657	Yellow	*H. doris*	*doris*	Cauca Valley	0.00	0.00	1.10	0.00	0.00	2.57	1.13	0.00	4.81
5012	White_yellow	*H. eleuchia*	*eleuchia*	Caldas	0.00	0.00	0.00	0.00	0.00	1.37	0.95	0.68	2.99
5013	White_yellow	*H. eleuchia*	*eleuchia*	Caldas	0.00	0.00	0.00	0.00	3.68	0.51	0.27	0.00	4.46
5014	White_yellow	*H. eleuchia*	*eleuchia*	Caldas	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.11
5499	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	4.44	0.00	0.10	2.35	0.38	0.00	7.27
5500	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	0.00	0.15	72.69	0.00	0.30	0.00	73.14
5501	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	0.00	0.70	67.64	0.00	0.35	0.00	68.68
5502	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	0.00	0.09	86.58	0.11	0.40	0.00	87.18
5673	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	0.00	0.00	74.51	0.00	0.00	0.00	74.51
5754	White_yellow	*H. eleuchia*	*eleuchia*	Cauca Valley	0.00	0.00	0.00	0.00	78.61	0.00	0.00	0.00	78.61
4007	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas	0.00	0.00	0.00	0.00	0.00	3.31	1.58	0.00	4.89
4008	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas	0.00	0.00	1.32	0.00	0.00	9.61	4.69	1.88	17.50
5023	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas	0.00	0.00	0.91	0.53	0.00	9.58	1.67	1.11	13.80
5329	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas-Pedrera	0.00	0.00	0.51	0.00	0.00	14.36	9.31	4.96	29.14
5389	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	6.85	2.09	0.33	9.27
5514	Dennis-ray	*H. elevatus*	*elevatus*	Amazonas	0.00	0.00	5.41	0.00	0.00	12.61	1.63	0.00	19.65

5490	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	9.03	0.00	0.00	5.00	0.00	10.09	1.25	0.00	25.37
5491	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	21.98	0.68	0.00	2.13	0.00	3.60	1.05	0.00	29.43
5492	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	3.49	0.00	0.00	0.00	0.00	18.32	6.92	3.63	32.36
5493	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	1.88	0.00	0.00	5.30	0.49	11.16	4.20	2.99	26.01
5494	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	0.00	0.00	21.59	0.00	0.47	11.43	1.87	0.00	35.35
5715	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	0.00	0.00	0.00	1.92	0.00	25.16	4.03	0.00	31.12
5716	Yellow_hindwing	*H. erato*	*chestertonii*	Cauca Valley	0.00	0.00	0.00	14.54	0.00	9.53	2.07	0.00	26.14
4774	Postman	*H. erato*	*dignus*	Putumayo	0.00	0.00	0.00	3.73	0.00	0.12	0.00	0.00	3.85
4775	Postman	*H. erato*	*dignus*	Putumayo	0.00	0.00	0.00	1.24	0.00	0.31	0.00	0.15	1.70
4777	Postman	*H. erato*	*dignus*	Putumayo	0.00	0.00	0.00	0.97	0.00	0.03	0.00	1.99	3.00
5599	Red band	*H. erato*	*guarica*	Boyaca	0.00	0.00	14.41	9.12	0.00	13.97	4.53	0.00	42.04
5600	Red band	*H. erato*	*guarica*	Boyaca	0.00	0.00	0.00	11.57	0.00	3.28	1.23	3.05	19.13
5601	Red band	*H. erato*	*guarica*	Boyaca	0.00	0.00	7.58	6.94	0.00	3.94	0.49	0.00	18.95
5626	Red band	*H. erato*	*guarica*	Boyaca	0.00	0.00	0.00	0.00	0.00	6.00	2.81	0.41	9.22
5940	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	122.24	0.00	0.00	0.00	0.00	122.24
5941	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	195.50	0.00	0.00	0.00	0.00	195.50
5942	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	0.00	0.00	9.54	4.53	1.15	15.22
5949	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	0.00	0.00	48.14	17.03	10.64	75.81
5950	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	1.94	0.00	24.83	8.48	0.00	35.24
5951	Red band	*H. erato*	*guarica*	Cundinamarca	0.00	0.00	0.00	8.01	0.00	37.99	21.01	0.00	67.01
4887	Red band	*H. erato*	*hydara*	Meta	0.00	0.00	0.00	27.42	0.00	0.01	0.00	0.24	27.66
4889	Red band	*H. erato*	*hydara*	Meta	0.00	0.00	0.00	4.22	0.00	0.00	0.00	1.37	5.58
4890	Red band	*H. erato*	*hydara*	Meta	0.00	0.00	0.00	43.64	0.00	0.00	0.00	0.37	44.01
4891	Red band	*H. erato*	*hydara*	Meta	0.00	0.00	0.00	7.80	0.00	0.00	0.00	0.00	7.80
4899	Red band	*H. erato*	*hydara*	Guaviare	0.00	0.00	0.00	0.00	0.00	0.06	0.00	1.20	1.27
4902	Red band	*H. erato*	*hydara*	Guaviare	0.00	0.00	0.00	2.29	0.00	0.75	0.00	2.18	5.22
4904	Red band	*H. erato*	*hydara*	Guaviare	0.00	0.00	0.00	2.81	0.04	0.30	0.00	0.00	3.16
4910	Red band	*H. erato*	*hydara*	Guaviare	0.00	0.00	0.00	0.39	0.00	1.10	0.54	1.85	3.88
4911	Red band	*H. erato*	*hydara*	Guaviare	0.00	0.00	0.00	4.28	0.00	0.00	0.00	0.00	4.23

5020	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.54	0.87
5025	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	10.83	0.00	0.09	0.00	2.05	12.97
5031	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	15.28	0.00	0.00	0.00	1.29	16.57
5303	Dennis-ray	*H. erato*	*reductimacula*	Amazonas-Pedrera	0.00	0.00	0.00	0.67	0.14	0.37	0.26	0.50	1.94
5305	Dennis-ray	*H. erato*	*reductimacula*	Amazonas-Pedrera	0.00	0.00	0.00	0.40	0.00	0.00	0.00	1.37	1.76
5306	Dennis-ray	*H. erato*	*reductimacula*	Amazonas-Pedrera	0.00	0.00	0.00	4.49	0.00	0.21	0.00	2.60	7.31
5307	Dennis-ray	*H. erato*	*reductimacula*	Amazonas-Pedrera	0.00	0.00	0.00	8.19	0.13	0.00	0.00	0.43	8.75
5322	Dennis-ray	*H. erato*	*reductimacula*	Amazonas-Pedrera	0.00	0.00	0.00	3.27	0.00	0.05	0.00	1.78	5.09
5508	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.37	0.00	0.00	2.41	0.00	0.00	1.45	0.91	5.15
5515	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	16.02	0.00	0.00	0.00	0.87	16.89
5516	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	3.13	0.00	0.00	0.00	0.52	3.65
5526	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	28.38	0.00	0.00	0.00	1.69	30.07
5530	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	1.46	0.00	0.00	0.00	1.83	3.29
5536	Dennis-ray	*H. erato*	*reductimacula*	Amazonas	0.00	0.00	0.00	11.40	0.00	0.00	0.00	1.90	13.30
4728	Postman yellow_ventral	*H. erato*	*venus*	Buenaventura	0.00	0.00	0.00	0.65	1.26	24.24	15.01	0.00	41.17
4729	Postman yellow_ventral	*H. erato*	*venus*	Buenaventura	0.00	0.00	0.00	1.04	0.68	19.82	11.36	0.00	32.91
4731	Postman yellow_ventral	*H. erato*	*venus*	Buenaventura	0.00	0.00	0.00	0.00	0.66	9.97	5.89	3.39	19.91
4732	Postman yellow_ventral	*H. erato*	*venus*	Buenaventura	0.00	0.00	0.00	1.04	3.75	18.24	12.07	1.18	36.28
4733	Postman yellow_ventral	*H. erato*	*venus*	Buenaventura	0.00	0.00	0.00	0.64	1.91	25.64	14.10	3.72	46.02

4122	No mimicry ring	*H. hecale*	*holcophorus*	Amazonas	0.00	0.00	0.00	0.00	0.00	5.13	1.47	0.00	6.60
5557	No mimicry ring	*H. hecale*	*holcophorus*	Amazonas	0.00	0.00	0.00	0.00	0.00	4.18	2.82	1.40	8.41
4011	No mimicry ring	*H. hecale*	*melicerta*	Santander	0.00	0.00	0.00	0.00	0.00	5.42	2.45	0.00	7.87
4012	No mimicry ring	*H. hecale*	*melicerta*	Santander	0.00	0.00	0.00	0.00	0.00	2.94	1.87	0.00	4.81
4207	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	10.80	0.00	4.47	2.09	0.00	17.36
4238	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	7.84	0.00	5.75	2.49	0.00	16.09
4239	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	8.05	0.00	4.50	1.31	0.00	13.86
4241	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	0.05	0.00	12.82	5.62	0.00	18.49
4242	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	0.00	0.00	8.56	4.32	0.00	12.89
5615	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	0.00	0.00	15.47	5.87	0.00	21.34
5616	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	12.12	0.00	10.73	6.12	1.48	30.44
5617	No mimicry ring	*H. hecale*	*melicerta*	Boyaca	0.00	0.00	0.00	0.00	3.74	7.38	4.43	0.00	15.55
5787	Tiger	*H. hecale*	*melicerta*	Buenaventura	0.00	0.00	0.00	44.94	0.00	0.00	0.00	0.00	44.94
5016	No mimicry ring	*H. hecalesia*	*hecalesia*	Caldas	0.82	0.00	0.00	0.00	0.00	1.42	0.21	0.00	2.45
5017	No mimicry ring	*H. hecalesia*	*hecalesia*	Caldas	0.00	0.00	0.00	0.00	0.00	5.64	8.25	0.32	14.21
5637	No mimicry ring	*H. hecalesia*	*hecalesia*	Boyaca	0.00	0.00	0.00	13.39	0.00	12.36	7.20	1.69	34.64
5479	No mimicry ring	*H. heurippa*	*heurippa*	Meta	0.00	0.24	0.00	0.00	0.00	11.64	6.31	4.47	22.65
4236	No mimicry ring	*H. ismenius*	*ismenius*	Boyaca	0.00	0.00	9.56	0.00	0.00	5.11	2.85	0.00	17.52
4237	No mimicry ring	*H. ismenius*	*ismenius*	Boyaca	0.00	0.00	8.94	0.00	0.00	6.42	3.63	0.00	18.98
4575	Tiger	*H. ismenius*	*boulleti*	Buenaventura	0.00	0.00	18.29	0.00	0.00	8.40	5.23	0.00	31.91
4578	Tiger	*H. ismenius*	*boulleti*	Buenaventura	0.00	0.00	15.90	0.00	0.00	7.29	3.87	0.00	27.06
4579	Tiger	*H. ismenius*	*boulleti*	Buenaventura	0.00	0.00	22.40	0.00	0.00	5.51	2.77	0.00	30.68
4580	Tiger	*H. ismenius*	*boulleti*	Buenaventura	0.00	0.00	18.94	0.00	0.00	10.19	4.79	0.00	33.92
4581	Tiger	*H. ismenius*	*boulleti*	Buenaventura	0.00	0.00	19.62	0.00	0.17	6.52	3.41	1.45	31.16
5532	Yellow	*H. leucadia*	*leucadia*	Amazonas	0.00	0.00	0.00	0.00	0.00	0.00	8.43	0.00	8.43
5544	Yellow	*H. leucadia*	*leucadia*	Amazonas	0.00	0.00	0.00	65.01	0.00	0.72	0.80	0.00	66.53
5609	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	6.16	3.70	0.00	9.86
5610	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	10.77	4.91	0.00	15.68
5611	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	11.76	4.52	0.00	16.28

5633	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	11.74	4.45	0.00	16.19
5634	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	8.51	3.98	0.00	12.49
5635	Red band	*H. melpomene*	*martinae*	Boyaca	0.00	0.00	0.00	0.00	0.00	42.13	1.44	0.00	43.57
4888	Red band	*H. melpomene*	*melpomene*	Meta	0.00	0.00	0.00	0.00	24.11	11.48	4.68	0.00	40.27
4892	Red band	*H. melpomene*	*melpomene*	Meta	0.00	0.00	0.00	0.00	5.26	26.80	8.85	0.00	40.92
4900	Red band	*H. melpomene*	*melpomene*	Guaviare	0.00	2.48	0.00	0.00	0.00	16.84	5.28	0.58	25.18
4903	Red band	*H. melpomene*	*melpomene*	Guaviare	0.00	0.97	0.00	0.00	0.00	21.52	10.83	0.00	33.32
4905	Red band	*H. melpomene*	*melpomene*	Guaviare	0.00	0.00	0.00	0.00	0.00	9.78	3.42	0.21	13.41
4906	Red band	*H. melpomene*	*melpomene*	Guaviare	0.00	2.72	0.00	0.00	0.00	23.29	11.80	6.24	44.05
4996	Red band	*H. melpomene*	*melpomene*	Meta	0.00	0.00	0.00	0.00	2.03	21.64	11.02	1.28	35.97
4997	Red band	*H. melpomene*	*melpomene*	Meta	0.00	0.00	0.00	0.00	5.73	22.73	8.74	0.79	37.99
4998	Red band	*H. melpomene*	*melpomene*	Meta	0.00	0.00	1.39	0.00	3.56	27.83	14.94	0.00	47.72
4039	Dennis-ray	*H. melpomene*	*vicina*	Amazonas	0.00	0.00	0.00	0.00	0.00	11.93	3.76	0.93	16.62
5027	Dennis-ray	*H. melpomene*	*vicina*	Amazonas	0.00	0.00	0.00	0.00	0.00	15.25	6.62	0.00	21.87
5321	Dennis-ray	*H. melpomene*	*vicina*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	9.24	5.66	2.67	17.57
5338	Dennis-ray	*H. melpomene*	*vicina*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	7.13	2.64	0.00	9.77
5339	Dennis-ray	*H. melpomene*	*vicina*	Amazonas-Pedrera	0.00	0.00	1.38	0.00	0.00	12.59	4.54	2.65	21.17
5340	Dennis-ray	*H. melpomene*	*vicina*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	11.00	5.32	3.26	19.58
5374	Dennis-ray	*H. melpomene*	*vicina*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	12.92	6.75	4.04	23.70
5533	Dennis-ray	*H. melpomene*	*vicina*	Amazonas	0.00	0.00	0.00	0.00	0.00	16.61	5.39	0.00	22.00
5534	Dennis-ray	*H. melpomene*	*vicina*	Amazonas	0.00	0.00	0.00	0.00	0.00	20.01	9.73	0.00	29.74
5561	Dennis-ray	*H. melpomene*	*vicina*	Amazonas	0.00	0.00	0.00	0.00	0.00	20.07	5.46	0.00	25.53
4727	Postman yellow_ventral	*H. melpomene*	*vulcanus*	Buenaventura	3.09	0.00	0.00	0.00	0.00	22.03	7.25	0.00	32.37

4730	Postman yellow_ventral	*H. melpomene*	*vulcanus*	Buenaventura	0.00	0.00	0.00	0.00	0.00	21.75	9.39	0.00	31.14
4741	Postman yellow_ventral	*H. melpomene*	*vulcanus*	Buenaventura	0.00	0.00	0.00	0.00	0.00	14.20	6.11	0.23	20.54
4743	Postman yellow_ventral	*H. melpomene*	*vulcanus*	Buenaventura	5.72	0.00	0.00	0.00	0.00	14.19	6.88	0.31	27.09
4744	Postman yellow_ventral	*H. melpomene*	*vulcanus*	Buenaventura	0.00	0.00	0.00	0.00	0.00	17.34	7.63	0.41	25.38
4778	Postman	*H. melpomene*	*bellula*	Putumayo	0.00	1.43	0.00	0.00	0.00	5.56	1.29	0.00	8.27
4779	Postman	*H. melpomene*	*bellula*	Putumayo	0.00	1.28	0.00	0.00	0.00	6.70	1.23	0.00	9.21
5388	No mimicry ring	*H. numata*	*silvana*	Amazonas-Pedrera	0.00	0.00	0.00	8.48	0.00	4.64	2.49	1.85	17.47
5392	No mimicry ring	*H. numata*	*silvana*	Amazonas-Pedrera	0.00	0.00	0.00	3.33	0.00	7.48	5.21	4.37	20.39
5422	No mimicry ring	*H. numata*	*silvana*	Amazonas-Pedrera	0.00	0.00	0.00	4.93	0.00	12.06	5.99	4.50	27.47
5423	No mimicry ring	*H. numata*	*silvana*	Amazonas-Pedrera	0.00	0.00	0.00	36.10	0.00	0.65	0.15	0.00	36.90
5424	No mimicry ring	*H. numata*	*silvana*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	3.88	1.76	3.17	8.81
5509	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	16.95	0.00	0.00	7.52	0.77	0.24	25.48
5510	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	5.04	0.00	0.00	10.21	4.84	0.15	20.24
5511	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.49	0.00	10.09	4.46	1.66	0.69	17.39
5512	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.00	0.00	0.00	15.08	5.32	0.47	20.86
5513	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.00	0.00	0.00	14.59	4.62	0.32	19.53
5519	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.00	0.00	0.00	12.94	4.90	0.32	18.16
5523	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.00	0.00	0.66	11.76	8.59	5.21	26.22
5548	No mimicry ring	*H. numata*	*silvana*	Amazonas	0.00	0.00	0.00	0.00	0.00	15.35	8.27	3.48	27.10
4006	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas	0.00	0.00	0.00	0.00	0.00	8.21	2.64	0.00	10.85

5330	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	15.91	6.41	0.89	23.20
5331	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	14.11	4.20	0.59	18.91
5332	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	10.59	3.95	0.41	14.96
5333	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.93	18.01	7.14	0.37	26.44
5520	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas	0.00	0.00	0.00	0.00	0.00	17.51	7.83	1.20	26.53
5550	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas	0.00	0.00	0.00	0.00	0.00	13.81	7.44	2.00	23.25
5551	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas	0.00	0.00	0.00	0.00	0.00	35.71	21.11	11.02	67.84
5552	No mimicry ring	*H. pardalinus*	*butleri*	Amazonas	0.00	0.00	0.00	0.00	0.00	5.61	2.38	0.15	8.15
5786	White	*H. sapho*	*chocoensis*	Buenaventura	0.00	0.00	0.00	160.80	0.00	2.43	0.38	0.00	163.61
5788	White	*H. sapho*	*chocoensis*	Buenaventura	0.00	0.00	0.00	0.00	0.00	2.21	0.67	0.82	3.70
5790	White	*H. sapho*	*chocoensis*	Buenaventura	0.00	0.00	0.00	0.00	0.00	2.52	0.80	0.00	3.32
4183	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	25.43	0.00	10.96	5.50	0.00	41.89
4187	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	5.29	0.00	14.72	5.77	0.00	25.79
4231	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	32.24	0.00	1.07	0.68	0.00	33.99
4232	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	50.60	0.00	0.49	0.39	0.00	51.47
4233	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	18.20	0.00	1.38	0.22	0.00	19.81
4374	Yellow	*H. sara*	*magdalena*	Arauca	0.00	0.00	0.00	0.00	7.09	6.18	3.93	1.11	18.31
4375	Yellow	*H. sara*	*magdalena*	Arauca	0.00	0.00	0.00	87.13	0.09	0.01	0.04	0.00	87.28
4376	Yellow	*H. sara*	*magdalena*	Arauca	0.00	0.00	0.00	0.01	20.05	7.75	3.61	0.00	31.42
4377	Yellow	*H. sara*	*magdalena*	Arauca	0.00	0.00	0.00	86.56	0.00	0.05	0.00	0.00	86.61
4673	Yellow	*H. sara*	*magdalena*	Cauca Valley	0.00	0.00	0.00	42.68	0.00	26.44	13.45	0.00	82.57
4674	Yellow	*H. sara*	*magdalena*	Cauca Valley	0.00	0.00	0.00	5.50	0.00	21.64	10.86	2.17	40.16
4683	Yellow	*H. sara*	*magdalena*	Cauca Valley	0.00	0.00	0.00	3.91	0.00	23.92	15.55	10.36	53.73
4684	Yellow	*H. sara*	*magdalena*	Cauca Valley	0.00	0.00	0.00	0.73	0.00	4.91	3.15	1.33	10.12
5571	Yellow	*H. sara*	*magdalena*	Meta	0.00	0.00	0.00	49.94	0.00	3.30	2.35	0.00	55.60

5572	Yellow	*H. sara*	*magdalena*	Meta	0.00	0.00	0.00	56.44	0.00	5.91	3.85	0.00	66.20
5573	Yellow	*H. sara*	*magdalena*	Meta	0.00	0.00	0.00	49.74	0.00	4.12	3.56	0.00	57.39
5574	Yellow	*H. sara*	*magdalena*	Meta	0.00	0.00	0.00	58.21	0.00	6.08	2.02	0.00	66.32
5575	Yellow	*H. sara*	*magdalena*	Meta	0.00	0.00	0.00	22.36	0.00	14.80	5.55	2.62	45.33
5650	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	0.75	0.00	5.98	3.55	1.22	11.49
5651	Yellow	*H. sara*	*magdalena*	Boyaca	0.00	0.00	0.00	0.00	0.00	13.80	4.59	0.00	18.39
5762	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	47.99	0.00	44.93	25.21	8.21	126.33
5763	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	70.49	0.00	8.92	6.55	1.27	87.22
5764	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	75.79	0.00	59.24	36.43	7.95	179.41
5774	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	84.54	0.00	7.25	3.67	0.00	95.46
5775	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	44.69	0.00	4.71	3.39	0.00	52.79
5776	Yellow	*H. sara*	*magdalena*	Buenaventura	0.00	0.00	0.00	76.94	0.00	4.71	3.59	0.00	85.24
4347	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	37.97	0.00	5.48	3.84	0.42	47.72
4348	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	53.28	0.00	11.86	7.11	0.69	72.94
4349	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	0.65	42.06	1.53	0.91	0.07	45.21
4350	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	7.63	0.00	13.67	7.99	1.30	30.59
4351	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	0.00	29.26	2.46	0.84	0.09	32.64
4352	Yellow	*H. sara*	*magdalena*	Vichada	0.00	0.00	0.00	0.00	61.68	3.91	2.27	0.12	67.98
4947	Yellow	*H. sara*	*sara*	Guaviare	0.00	0.00	0.00	65.96	0.00	8.47	3.14	0.00	77.57
5527	Yellow	*H. sara*	*sara*	Amazonas	0.00	0.00	0.00	24.53	0.00	0.00	6.13	1.46	32.12
5538	Yellow	*H. sara*	*sara*	Amazonas	0.00	0.00	0.00	0.00	0.00	9.09	5.84	0.42	15.35
5540	Yellow	*H. sara*	*sara*	Amazonas	0.00	0.00	0.00	0.00	0.00	0.00	39.56	16.26	55.81
5541	Yellow	*H. sara*	*sara*	Amazonas	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.85
5542	Yellow	*H. sara*	*sara*	Amazonas	0.00	0.00	0.00	0.00	0.73	21.41	19.25	4.25	45.64
5308	No mimicry ring	*H. wallacei*	*flavescens*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	16.86	3.92	0.00	20.77
5310	No mimicry ring	*H. wallacei*	*flavescens*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	20.97	2.60	0.00	23.57

5320	No mimicry ring	*H. wallacei*	*flavescens*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	17.21	3.39	0.00	20.60
5342	No mimicry ring	*H. wallacei*	*flavescens*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.00	19.17	4.65	0.00	23.82
5343	No mimicry ring	*H. wallacei*	*flavescens*	Amazonas-Pedrera	0.00	0.00	0.00	0.00	0.93	22.54	12.76	10.94	47.17

Table S3.2 CNglcs identified in this study.

Cyanogenic glucoside	Source	Retention time (min)	Sodium adducts (m/z)	Formic adducts (m/z)
Linamarin	Biosynthesized	2.8	270	292
Lotaustralin	Biosynthesized	4	284	306
Epilotasutralin	Biosynthesized	4.3	284	
Deidaclin/Tetraphylin A	Sequestered	4	290	316
Tetraphylin B	Sequestered	1.4	310	332
Epivolkenin	Sequestered	2	310	332
Gynocardin	Sequestered	1	326	348
Dihydrogynocardin	Sequestered	1.8	328	350
Amygdalin	External Standard	5.3	480	502

Table S3.3 CNglc profiles compared among all species tested, between all species from the same locality, and between populations of the same species (PERMANOVA).

Comparison between	Sum of squares	R^2	F-value	p (>F)
All Species of *Heliconius*	66818	0.28	4.57	0.001***
Species occurring in Puerto Nariño-Amazonas	13289	0.53	4.17	0.003***
Species occurring in Pedrera-Amazonas	3584.2	0.35	2.14	0.034*
Species occurring in Boyaca	6482.6	0.53	4.84	0.001***
Species occurring in Buenaventura	27181	0.54	3.94	0.012*
Species occurring in Caldas	59.486	0.21	0.83	0.541
Species occurring in Guaviare	4611.6	0.95	76.99	0.001***
Species occurring in Meta	7751.1	0.75	11.36	0.002***
Species occurring in Putumayo	52.288	0.87	20.13	0.100
Species occurring in Valle del Cauca	23116	0.67	13.98	0.001***
Populations of *H. sara*	22826	0.46	3.69	0.001***
Populations of *H. erato*	17981	0.3	2.11	0.028*
Populations of *H. melpomene*	1066.1	0.36	2.3962	0.015*
Populations of *H. eleuchia*	7721.2	0.6	10.73	0.026*
Populations of *H. cydno*	655.34	0.63	8.76	0.007**

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S3.4 CNglc profiles compared between species pairs (PERMANOVA).

Species pairs	Sum of squares	F-values	R2	p-value	p-adjusted
H. antiochus vs. H. cydno	4415.52	8.71	0.26	0	0.01*
H. antiochus vs. H. eleuchia	12494.56	8.78	0.34	0	0.01*
H. antiochus vs. H. melpomene	5665.73	15.78	0.28	0	0.01*
H. charitonia vs. H. clysonimus	1011.41	3.52	0.2	0	0.01*
H. charitonia vs. H. cydno	983.88	7.95	0.26	0	0.01*
H. charitonia vs. H. hecale	911.54	4.6	0.2	0	0.01*
H. charitonia vs. H. melpomene	1880.01	15.75	0.29	0	0.01*
H. charitonia vs. H. numata	934.96	4.95	0.21	0	0.01*
H. charitonia vs. H. pardalinus	1254.42	7.32	0.33	0	0.01*
H. clysonimus vs. H. melpomene	1316.28	9.39	0.2	0	0.01*
H. cydno vs. H. ismenius	1372.52	21.34	0.49	0	0.01*
H. doris vs. H. eleuchia	8792.45	7.48	0.37	0	0.01*
H. eleuchia vs. H. erato	15183.96	11.57	0.17	0	0.01*
H. eleuchia vs. H. melpomene	13962.38	34.74	0.47	0	0.01*
H. eleuchia vs. H. pardalinus	9377.04	10.94	0.41	0	0.01*
H. eleuchia vs. H. sara	18780.07	13.62	0.23	0	0.01*
H. elevatus vs. H. ismenius	737.29	23.17	0.68	0	0.01*
H. erato vs. H. melpomene	4701.77	5.88	0.07	0	0.01*
H. hecale vs. H. ismenius	1391.21	10.75	0.37	0	0.01*
H. hecale vs. H. melpomene	1276.65	10.84	0.2	0	0.01*
H. ismenius vs. H. melpomene	1999.6	23.83	0.39	0	0.01*
H. ismenius vs. H. numata	1071.8	8.93	0.33	0	0.01*
H. ismenius vs. H. pardalinus	1369.39	16.82	0.55	0	0.01*
H. ismenius vs. H. sara	7546.49	6.56	0.13	0	0.01*
H. melpomene vs. H. sapho	8478.06	13.88	0.3	0	0.01*
H. melpomene vs. H. sara	17721.89	23.08	0.25	0	0.01*
H. antiochus vs. H. charitonia	3990.91	4.89	0.23	0.01	0.02*
H. antiochus vs. H. ismenius	4072.84	5.25	0.26	0.01	0.04*
H. antiochus vs. H. pardalinus	3478.46	4.77	0.22	0.02	0.04*
H. aoede vs. H. cydno	6347.16	81.08	0.83	0.01	0.03*
H. aoede vs. H. eleuchia	8573.75	6	0.4	0.02	0.05*
H. aoede vs. H. hecale	4810.32	27.84	0.68	0.01	0.03*
H. aoede vs. H. ismenius	5851.8	143.88	0.95	0.02	0.04*

H. aoede vs. H. melpomene	7021.45	74.39	0.7	0	0.01*
H. aoede vs. H. numata	5347.55	33.47	0.72	0.01	0.03*
H. aoede vs. H. pardalinus	6098.22	52.01	0.85	0.02	0.04*
H. charitonia vs. H. doris	2021.33	5.85	0.33	0	0.01*
H. charitonia vs. H. eleuchia	8324.13	8.69	0.37	0.01	0.03*
H. charitonia vs. H. elevatus	575.72	3.88	0.24	0.01	0.02*
H. charitonia vs. H. ismenius	1502.68	10.85	0.45	0	0.01*
H. charitonia vs. H. sara	7863.8	6.8	0.13	0.01	0.02*
H. charitonia vs. H. wallacei	1196.28	7.29	0.4	0	0.01*
H. clysonimus vs. H. cydno	674.03	4.26	0.16	0.01	0.04*
H. clysonimus vs. H. doris	1511.51	3.68	0.23	0.01	0.04*
H. clysonimus vs. H. eleuchia	8342.3	8.25	0.35	0.01	0.04*
H. clysonimus vs. H. ismenius	1187.75	5.96	0.31	0	0.01*
H. clysonimus vs. H. pardalinus	861.65	3.85	0.2	0	0.01*
H. cydno vs. H. doris	981.05	5.48	0.21	0.02	0.04*
H. cydno vs. H. eleuchia	10811.05	18.55	0.44	0	0.01*
H. cydno vs. H. hecale	422.17	3.5	0.11	0.01	0.04*
H. cydno vs. H. leucadia	2059.15	10.38	0.38	0.02	0.04*
H. cydno vs. H. melpomene	445.69	5.05	0.1	0.02	0.05*
H. cydno vs. H. sara	11567.34	12.13	0.19	0	0.01*
H. doris vs. H. hecale	1546.11	5.62	0.25	0.01	0.02*
H. doris vs. H. ismenius	1858.66	7.54	0.41	0	0.01*
H. doris vs. H. numata	1086.79	4.1	0.19	0.01	0.04*
H. doris vs. H. sapho	6816.46	2.41	0.26	0.02	0.04*
H. doris vs. H. sara	6234.68	5.05	0.11	0.01	0.02*
H. eleuchia vs. H. elevatus	6878.26	6.92	0.35	0.02	0.04*
H. eleuchia vs. H. hecale	10006.32	13.43	0.4	0	0.01*
H. eleuchia vs. H. ismenius	8347.38	9.03	0.39	0.01	0.02*
H. eleuchia vs. H. numata	9901.76	13.44	0.4	0	0.01*
H. eleuchia vs. H. wallacei	7040.09	6.53	0.35	0.02	0.05*
H. elevatus vs. H. sapho	5847.71	2.35	0.25	0.01	0.04*
H. elevatus vs. H. wallacei	288.23	7.32	0.45	0.01	0.02*
H. erato vs. H. sara	8465.54	6.54	0.07	0.01	0.02*
H. hecale vs. H. pardalinus	724.53	4.67	0.19	0	0.02*
H. hecale vs. H. sara	6361.06	6.06	0.11	0.01	0.03*
H. hecale vs. H. wallacei	759.61	5.21	0.25	0.01	0.02*
H. hecalesia vs. H. ismenius	601.01	11.87	0.6	0.01	0.02*

H. hecalesia vs. H. wallacei	352.88	5.17	0.46	0.02	0.04*
H. ismenius vs. H. sapho	6653.89	3.06	0.28	0.01	0.03*
H. ismenius vs. H. wallacei	1228.84	32.92	0.77	0	0.01*
H. leucadia vs. H. melpomene	2457.43	15.53	0.33	0	0.01*
H. leucadia vs. H. pardalinus	2118.86	6.16	0.41	0.02	0.05*
H. melpomene vs. H. numata	646.08	5.67	0.12	0	0.02*
H. numata vs. H. sara	7363.05	7.04	0.13	0	0.02*
H. numata vs. H. wallacei	444.32	3.28	0.17	0.02	0.05*
H. pardalinus vs. H. sapho	6948.47	3.82	0.28	0.01	0.02*
H. pardalinus vs. H. sara	7432.37	6.65	0.13	0	0.01*
H. sapho vs. H. wallacei	5977.33	2.06	0.26	0.02	0.05*
H. antiochus vs. H. aoede	1959.76	1.7	0.15	0.2	0.28
H. antiochus vs. H. clysonimus	1286.14	1.49	0.08	0.23	0.31
H. antiochus vs. H. doris	3440.42	3.45	0.2	0.06	0.11
H. antiochus vs. H. elevatus	2658.54	3.21	0.19	0.08	0.13
H. antiochus vs. H. erato	1850.86	1.46	0.03	0.19	0.27
H. antiochus vs. H. hecale	2431.63	3.76	0.15	0.05	0.09
H. antiochus vs. H. hecalesia	1170.43	1.1	0.09	0.32	0.41
H. antiochus vs. H. heurippa	643.27	0.51	0.05	0.63	0.71
H. antiochus vs. H. leucadia	241.08	0.18	0.02	0.86	0.87
H. antiochus vs. H. numata	2895.47	4.53	0.18	0.03	0.05*
H. antiochus vs. H. sapho	1912.13	0.73	0.06	0.48	0.57
H. antiochus vs. H. sara	340.34	0.26	0.01	0.83	0.86
H. antiochus vs. H. wallacei	2623.09	2.93	0.18	0.11	0.17
H. aoede vs. H. charitonia	5736.49	26.75	0.77	0.03	0.06
H. aoede vs. H. clysonimus	3816.56	12.18	0.6	0.02	0.05*
H. aoede vs. H. doris	6133.24	14.01	0.7	0.03	0.05*
H. aoede vs. H. elevatus	5291.53	119.43	0.95	0.04	0.08
H. aoede vs. H. erato	4226.71	3.41	0.07	0.07	0.13
H. aoede vs. H. hecalesia	3600.55	33.66	0.92	0.1	0.16
H. aoede vs. H. heurippa	2419.54	24.13	0.96	0.33	0.42
H. aoede vs. H. leucadia	702.38	0.63	0.24	0.67	0.74
H. aoede vs. H. sapho	37.38	0.01	0	1	1
H. aoede vs. H. sara	1747.58	1.34	0.03	0.32	0.41
H. aoede vs. H. wallacei	5490.09	94.96	0.95	0.05	0.09
H. charitonia vs. H. erato	1921.42	1.7	0.03	0.15	0.23
H. charitonia vs. H. hecalesia	357.78	1.75	0.16	0.19	0.27

H. charitonia vs. H. heurippa	203.27	0.88	0.11	0.21	0.29
H. charitonia vs. H. leucadia	1933.6	4.12	0.34	0.04	0.09
H. charitonia vs. H. sapho	6566.19	3.13	0.26	0.04	0.07
H. clysonimus vs. H. elevatus	387.25	1.81	0.13	0.24	0.32
H. clysonimus vs. H. erato	38.28	0.03	0	0.99	1
H. clysonimus vs. H. hecale	94.12	0.39	0.02	0.53	0.62
H. clysonimus vs. H. hecalesia	97.05	0.33	0.04	0.78	0.83
H. clysonimus vs. H. heurippa	142.24	0.41	0.06	0.57	0.66
H. clysonimus vs. H. leucadia	864.59	1.52	0.16	0.16	0.24
H. clysonimus vs. H. numata	245.1	1.06	0.05	0.35	0.44
H. clysonimus vs. H. sapho	4139.6	1.9	0.17	0.22	0.31
H. clysonimus vs. H. sara	3195.57	2.72	0.06	0.09	0.15
H. clysonimus vs. H. wallacei	844.03	3.58	0.25	0.05	0.1
H. cydno vs. H. elevatus	20.74	0.31	0.01	0.71	0.78
H. cydno vs. H. erato	2053.62	2.13	0.03	0.1	0.16
H. cydno vs. H. hecalesia	93.87	1.16	0.06	0.31	0.4
H. cydno vs. H. heurippa	12.62	0.16	0.01	0.84	0.86
H. cydno vs. H. numata	142.53	1.24	0.04	0.25	0.33
H. cydno vs. H. pardalinus	221.67	2.43	0.09	0.12	0.19
H. cydno vs. H. sapho	7493.13	7.3	0.29	0.02	0.05*
H. cydno vs. H. wallacei	337.5	4.76	0.19	0.03	0.06
H. doris vs. H. elevatus	742.07	2.76	0.22	0.09	0.15
H. doris vs. H. erato	2554.54	2.14	0.04	0.15	0.23
H. doris vs. H. hecalesia	699.05	1.78	0.2	0.18	0.26
H. doris vs. H. heurippa	156.79	0.31	0.06	0.6	0.68
H. doris vs. H. leucadia	2467.78	3.17	0.35	0.03	0.06
H. doris vs. H. melpomene	389.43	2.57	0.07	0.09	0.15
H. doris vs. H. pardalinus	307.14	1.15	0.08	0.35	0.44
H. doris vs. H. wallacei	102.01	0.34	0.04	0.74	0.8
H. eleuchia vs. H. hecalesia	4271.32	3.29	0.25	0.07	0.12
H. eleuchia vs. H. heurippa	1799.06	1.13	0.12	0.28	0.37
H. eleuchia vs. H. leucadia	4723.49	2.85	0.24	0.08	0.13
H. eleuchia vs. H. sapho	10537.03	3.51	0.26	0.03	0.06
H. elevatus vs. H. erato	835.73	0.73	0.01	0.28	0.36
H. elevatus vs. H. hecale	209.08	1.54	0.08	0.23	0.31
H. elevatus vs. H. hecalesia	66.06	1.2	0.15	0.27	0.35
H. elevatus vs. H. heurippa	20.97	0.63	0.11	0.55	0.63

H. elevatus vs. H. leucadia	1705.99	4.43	0.43	0.03	0.07
H. elevatus vs. H. melpomene	275.99	3.22	0.08	0.05	0.09
H. elevatus vs. H. numata	71.49	0.57	0.03	0.72	0.78
H. elevatus vs. H. pardalinus	186.4	2.16	0.14	0.13	0.21
H. elevatus vs. H. sara	5188.44	4.41	0.09	0.02	0.05*
H. erato vs. H. hecale	381.1	0.37	0.01	0.72	0.78
H. erato vs. H. hecalesia	194.33	0.16	0	0.78	0.83
H. erato vs. H. heurippa	192.67	0.15	0	0.46	0.56
H. erato vs. H. ismenius	2379.47	2.12	0.04	0.14	0.21
H. erato vs. H. leucadia	877.4	0.68	0.01	0.16	0.24
H. erato vs. H. numata	788.79	0.76	0.01	0.45	0.55
H. erato vs. H. pardalinus	1832.77	1.67	0.03	0.17	0.25
H. erato vs. H. sapho	4890.34	3.13	0.06	0.14	0.22
H. erato vs. H. wallacei	1413.81	1.21	0.02	0.16	0.24
H. hecale vs. H. hecalesia	21.12	0.12	0.01	0.85	0.87
H. hecale vs. H. heurippa	87.27	0.49	0.04	0.2	0.28
H. hecale vs. H. leucadia	1251.12	3.79	0.23	0.05	0.1
H. hecale vs. H. numata	126.42	0.74	0.03	0.46	0.56
H. hecale vs. H. sapho	5481.27	3.96	0.22	0.11	0.17
H. hecalesia vs. H. heurippa	46.78	0.42	0.17	0.75	0.81
H. hecalesia vs. H. leucadia	989.52	1.26	0.3	0.3	0.39
H. hecalesia vs. H. melpomene	310.04	3.26	0.09	0.05	0.09
H. hecalesia vs. H. numata	34.97	0.22	0.02	0.9	0.9
H. hecalesia vs. H. pardalinus	238.55	2.03	0.17	0.17	0.25
H. hecalesia vs. H. sapho	3678.8	0.84	0.17	0.7	0.77
H. hecalesia vs. H. sara	2086.25	1.64	0.04	0.21	0.29
H. heurippa vs. H. ismenius	270.48	8.8	0.59	0.11	0.18
H. heurippa vs. H. leucadia	804.48	0.38	0.27	0.67	0.74
H. heurippa vs. H. melpomene	33.47	0.36	0.01	0.6	0.68
H. heurippa vs. H. numata	34.24	0.21	0.02	0.79	0.83
H. heurippa vs. H. pardalinus	20.08	0.17	0.02	0.5	0.59
H. heurippa vs. H. sapho	2256.5	0.26	0.12	0.5	0.59
H. heurippa vs. H. sara	977.86	0.73	0.02	0.48	0.57
H. heurippa vs. H. wallacei	54.55	1.16	0.22	0.33	0.42
H. ismenius vs. H. leucadia	2124.61	6.39	0.48	0.03	0.05*
H. leucadia vs. H. numata	1552.85	4.9	0.27	0.05	0.09
H. leucadia vs. H. sapho	558.2	0.09	0.03	0.8	0.84

H. leucadia vs. H. sara	230.49	0.17	0	0.82	0.85
H. leucadia vs. H. wallacei	2032.48	4.36	0.47	0.05	0.09
H. melpomene vs. H. pardalinus	25.73	0.26	0.01	0.8	0.84
H. melpomene vs. H. wallacei	66.18	0.75	0.02	0.47	0.57
H. numata vs. H. pardalinus	354.3	2.42	0.11	0.05	0.09
H. numata vs. H. sapho	6143.51	4.48	0.24	0.1	0.16
H. pardalinus vs. H. wallacei	55.81	0.59	0.05	0.54	0.62
H. sapho vs. H. sara	1723.11	1.01	0.03	0.36	0.44
H. sara vs. H. wallacei	4773.31	3.95	0.09	0.03	0.05*

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

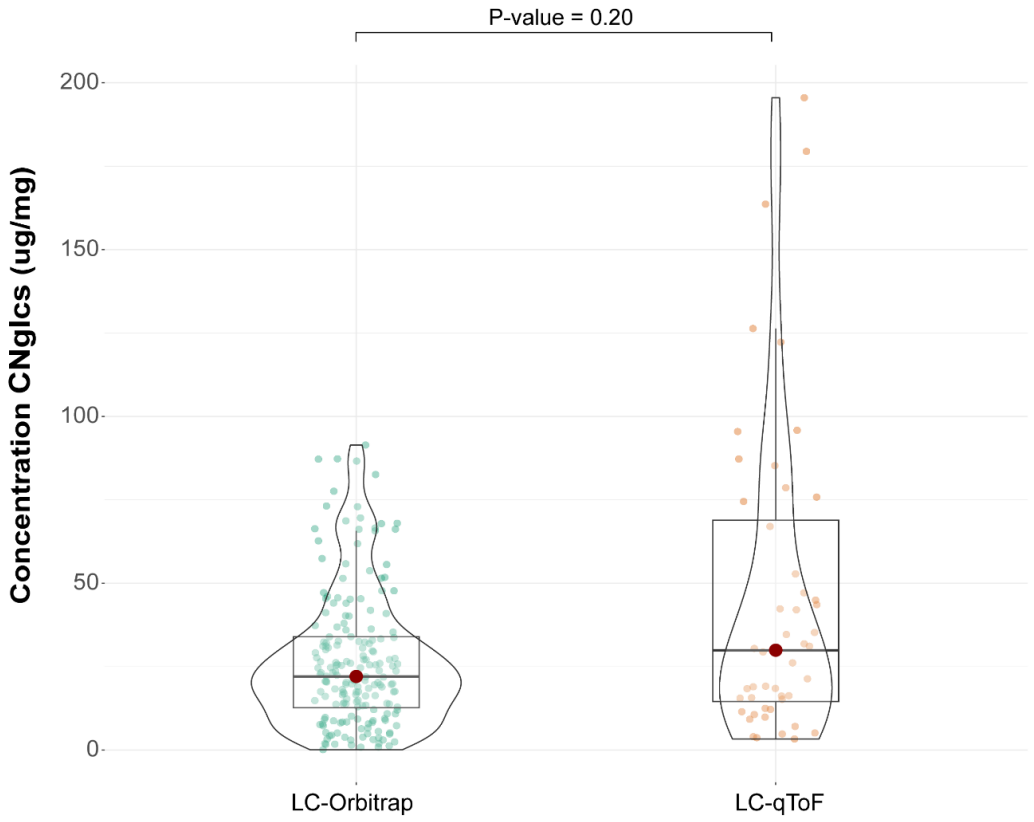

Figure S3.1 Differences between the quantification of all CNglcs combined ('CNglcs total') obtained from LC-qToF-MS and LC-Orbitrap-MS.

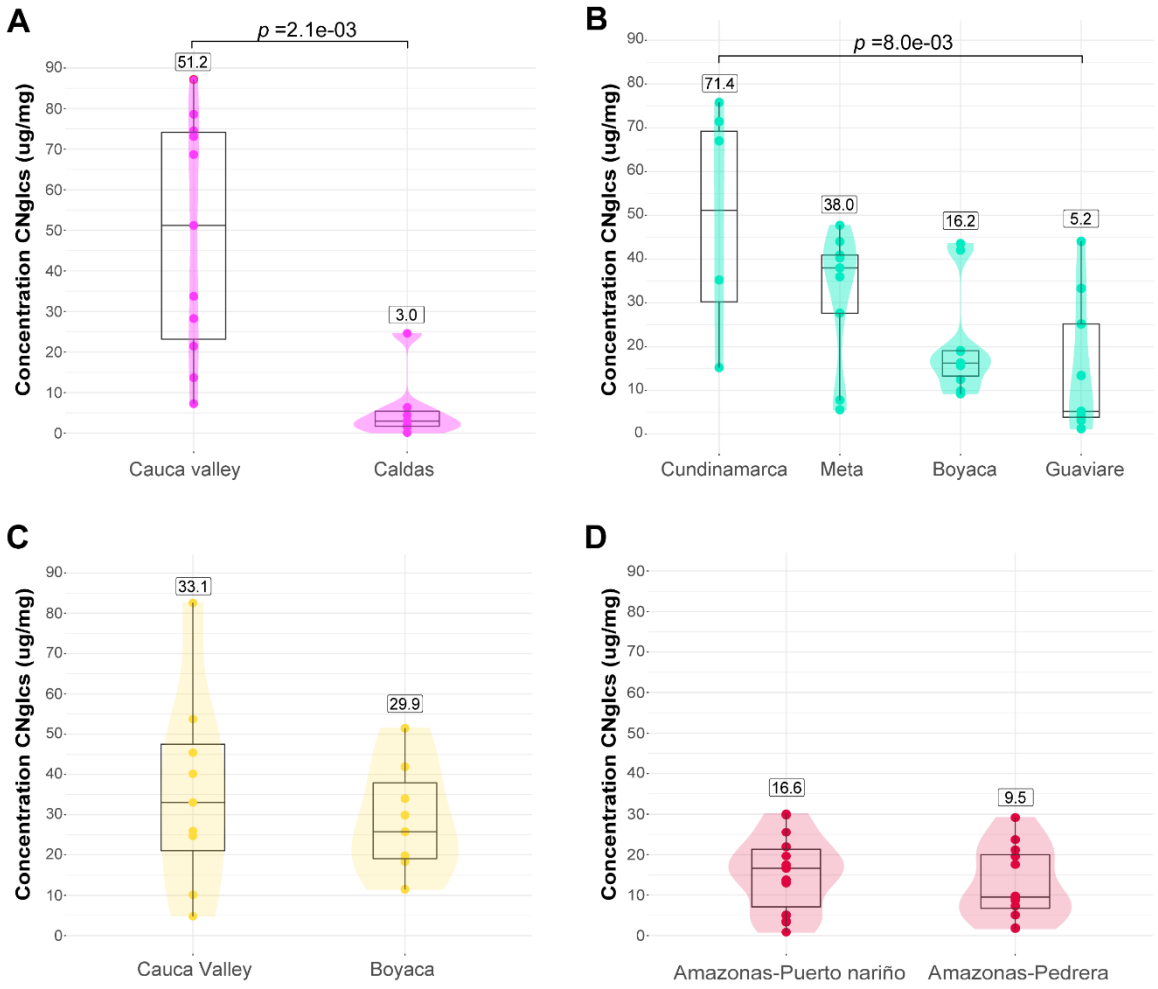

Figure S3.2 Quantification of all CNglcs combined ('CNglcs total') per mimicry ring and locality. (A) white yellow, (B) red band, (C) dennis-ray and (D) yellow. Mean concentration of *'CNglcs total'* is shown on top of each box (numbers on squares). Significantly different comparisons ($\alpha < 0.05$) are shown.

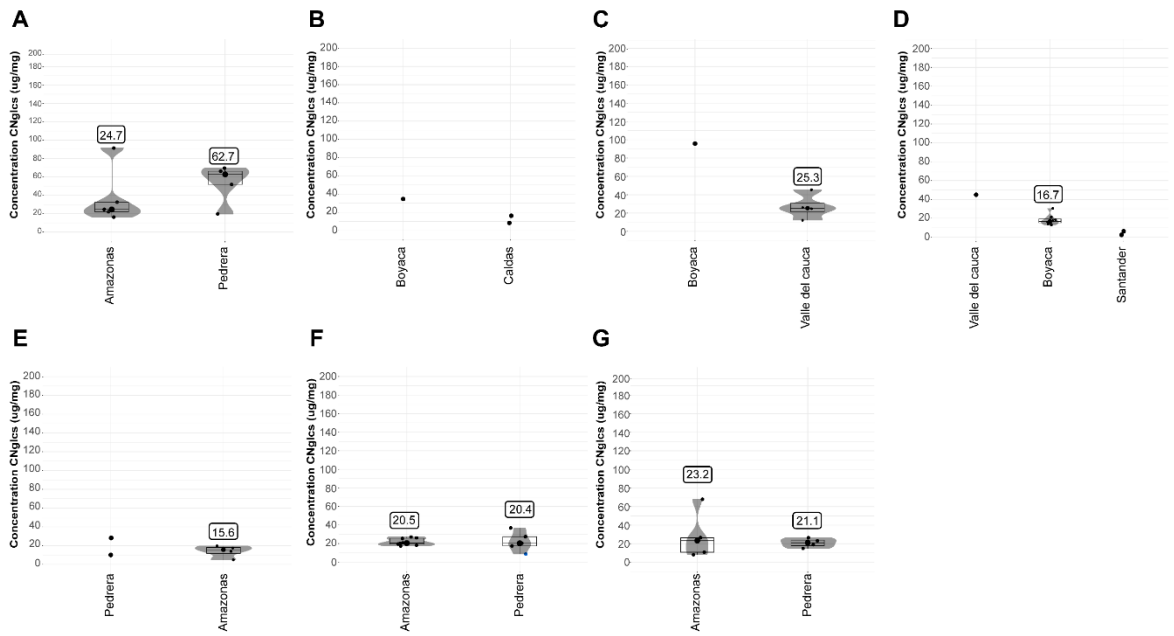

Figure S3.3 Variation in the concentrations of all CNglcs combined ('CNglcs total') within a species ocurring in different localities. (A) *H. antiochus*, (B) *H. hecalesia*, (C) *H. doris*, (D) *H. hecale*, (E) *H. elevatus,* (F) *H. numata,* and (G) *H. pardalinus*. Mean concentration of *'CNglcs total'* is shown on top of each box (numbers on squares).

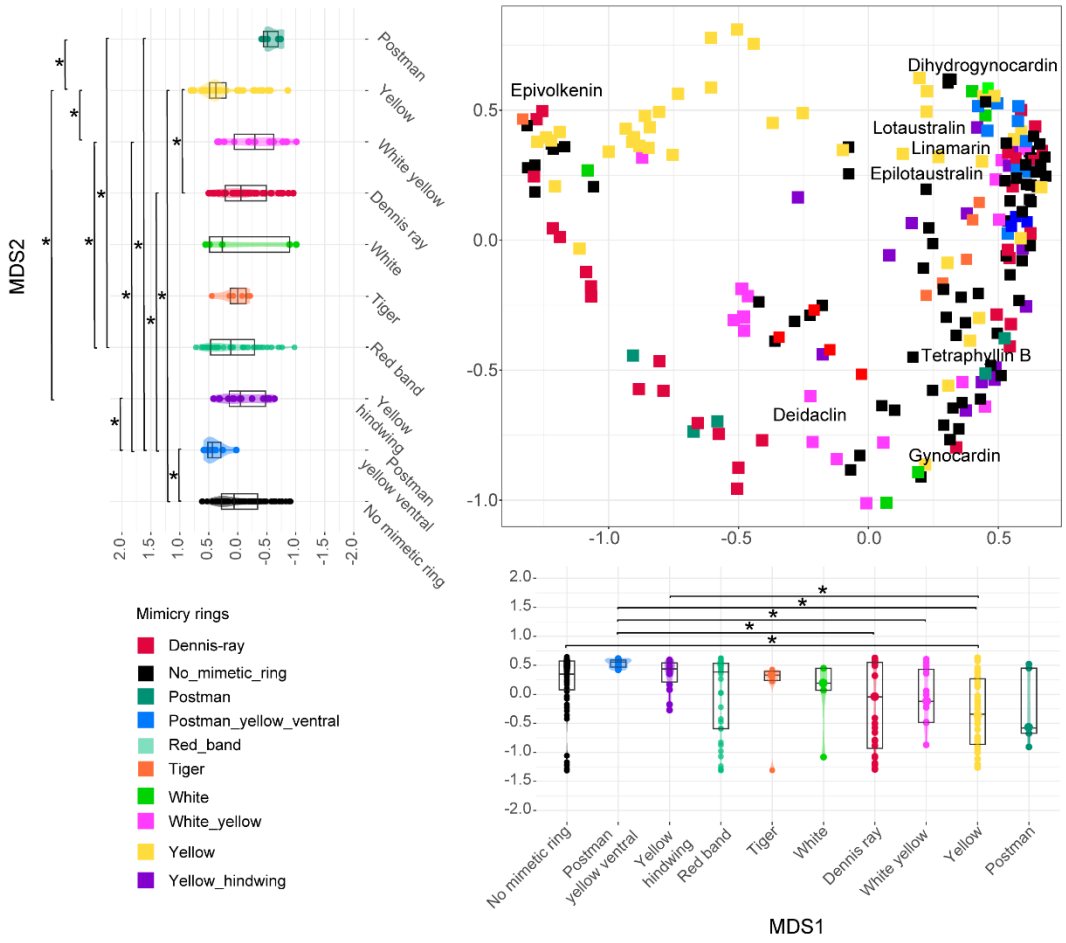

Figure S3.4 Non-metric multidimensional scaling (NMDS) of cyanogenic glycosides in *Heliconius* comparing mimicry rings. The distance between two symbols is indicative of how different two individuals are in the composition of CNglcs. Boxplots for both NMDS1 and NMDS2 show differences between mimicry rings in each axis. Asterisk (*) symbolises a *p*-value < 0.05.