# Sex chromosome – autosome fusions in *Heliconius* butterflies

Nicol Rueda - M[1], Chris D. Jiggins[2], Carolina Pardo-Diaz[1], W. Owen McMillan[3], Shane McCarthy[4], Krzysztof M. Kozak[3,7], Carlos F. Arias[3,8], Jonathan Ready[5,6], Richard Durbin[2,4], Gabriela Montejo-Kovacevich[2], Joana I. Meier[2,4*], Camilo Salazar[1*]✉

* These authors contributed equally to this work.

[1]Biology Program, Faculty of Natural Sciences, Universidad del Rosario, Bogotá, Colombia. 111221

[2]Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom

[3]Smithsonian Tropical Research Institute, Panama

[4]Tree of Life Programme, Welcome Sanger Institute, United Kingdom

[5]Centre for Advanced Studies of Biodiversity - CEABIO, Guamá Science Park, Lot 11, Avenida Perimetral da Ciência km 01 - Guamá, Belém - PA, 66075-750, Brazil

[6]Institute for Biological Sciences, Federal University of Pará - UFPA, R. Augusto Corrêa, 1 - Guamá, Belém - PA, 66075-110, Brazil

[7]Museum of Vertebrate Zoology, UC Berkeley, USA

[8]Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, United States

✉Correspondence: Camilo Salazar (camilo.salazar@urosario.edu.co)

## Abstract

Sex chromosomes are evolutionarily labile in many animals, and fusion with an autosome is one of the mechanisms by which they can evolve. Sex chromosome-autosome (SA) fusions can reduce sexual conflict and promote adaptation and reproductive isolation among species. Recently, advances in genomics have fuelled the discovery of SA fusions in multiple vertebrates and in some invertebrates such as Lepidoptera. Here, we discovered multiple SA fusions in the *sara/sapho* clade of the classical adaptive radiation of *Heliconius* butterflies. While the vast majority of *Heliconius* species have 21 chromosomes, species of this particularly rapidly diversifying clade have up to 60 chromosomes. We found a sex chromosome fusion with chromosome 4 at the base of the *sapho* clade and two additional sex chromosome fusions, each shared by two species. These sequential fusions between autosomes and sex chromosomes make the *Heliconius sapho* clade an ideal system to study the role of neo-sex chromosomes in adaptive radiations and the degeneration of sex chromosomes over time. Our study adds to a growing number of examples in butterflies that, in the future, may help to unravel the importance of such rearrangements in the evolution of Lepidoptera and animals in general.

**Keywords**

Sex chromosome-autosome fusions, *Heliconius*, sex chromosome evolution, adaptation.

## Introduction

Sex chromosome-autosome (SA) fusions contribute to the turnover of sex-determination loci or the evolution of neo-sex chromosomes (Ross et al. 2009; Yoshido et al. 2011; Šíchová et al. 2016), 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 (Kitano et al. 2009; Pennell et al. 2015; Sigeman et al. 2019; Jay et al. 2022). Sexually antagonistic selection is thought to favour the fusion of sex chromosomes with autosomes harbouring genes under sexually antagonistic selection (Charlesworth and Charlesworth 1980). There is limited evidence for this theory e.g. in sticklebacks (Ross et al. 2009) and warblers (Pala et al. 2012). 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 (Jay et al. 2022). SA fusions can also become fixed due to meiotic drive, such as female meiotic drive elements on W/X-A fusions that preferentially end up in the egg instead of the polar bodies (Pardo-Manuel de Villena and Sapienza 2001).

As with other chromosomal rearrangements, SA fusions can suppress recombination and potentially strengthen reproductive isolation if they bring together barrier loci into regions with reduced recombination (Guerrero and Kirkpatrick 2014; Lucek 2018; Hora et al. 2019; de Vos et al. 2020). 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 (Kitano et al. 2009). SA fusions can also facilitate adaptation, such as the SA fusion in *Cydia pomonella* (Tortricidae), which linked two insecticide-resistance genes and genes involved in detoxifying plant metabolites (Nguyen et al. 2013).

Genomic studies have revealed that SA fusions have occurred multiple times across vertebrates (Schmid et al. 2003; Matsumoto and Kitano 2016; Sigeman et al. 2019), and in some invertebrates such as mygalomorph spiders (Král et al. 2013), *Drosophila* (Flores et al. 2008) and true bugs of the genus *Dysdercus* (Bressa et al. 2009). In Lepidoptera, examples of SA fusions include *Danaus* (Smith et al. 2016; Mongue et al. 2017) and *Leptidea* butterflies (Šíchová et al. 2015; Šíchová et al. 2016), among others (Yoshido et al. 2005; Nguyen et al. 2013; Picq et al. 2018; Carabajal Paladino et al. 2019). Compared to taxa with centromeres, the holocentric chromosomes of Lepidoptera may facilitate fusions, as they are less likely to cause segregation problems during cytokinesis and thus reduce hybrid fitness (Lucek et al. 2022). Nonetheless, butterflies and moths have remarkably constrained chromosome evolution (White 1977), with most species having a haploid chromosome number ranging between 28 and 32 except for a few groups that have experienced extensive fission and fusion events (Hill et al. 2019). One of such cases are the *Heliconius* butterflies, which mostly have 21 chromosomes due to ancestral fusions (Davey et al. 2016). 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 (Brown 1981). This clade comprises 12 species (Jiggins and Lamas 2016), and differs from other *Heliconius* by showing an inability to synthesise

cyanogens, leading to reliance on toxins sequestered from larval host plants (Beltrán et al. 2007; Engler-Chaouat and Gilbert 2007). A subclade of seven species within the *sara/sapho* clade shows particularly high diversification rates (Kozak et al., 2015) and high number of chromosomes (Brown 1981).

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 S1) 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 (Ruan and Li 2020) . 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 (Garrison and Marth 2012) . Next, the assembly was polished using BCFtools consensus v1.9 (Danecek et al. 2021) 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 (Chow et al. 2016) . Lastly, Dovetail Genomics Hi-C data was used to scaffold the assembly further using SALSA v2.2 (Ghurye et al. 2019), followed by another round of manual curation with gEVAL (Chow et al. 2016). 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 (Manni et al. 2021) 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 S1). 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 (Li 2013) with default parameters. We then used samtools v1.12 to sort and index the alignment files (Li et al. 2009). PCR-duplicate reads were identified and removed using Picard tools v2.9.2 (Broad 2019), and variant calling was conducted with Haplotype Caller (GATK, v3.7.0) in BP-resolution mode (DePristo et al. 2011). Then, samples were jointly genotyped using GATK's GenotypeGVCFs (DePristo et al. 2011). We used vcftools v0.1.14 (Danecek et al. 2011) 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 less than 1.5 times the mean to exclude paralogous regions (Meier et al. 2018). For this, we used the custom script *removeTooHighDepthSites.sh* from Joana Meier's Github repository (https://github.com/joanam). 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 (Stamatakis et al. 2008), with the GTRGAMMA model and 100 bootstrap replicates. We applied the same procedure to obtain ML trees for each chromosome. In all cases, *H. charitonia* 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 (Danecek et al. 2011). To reduce the linkage disequilibrium effect, we used the python script *ldPruning.sh* (https://github.com/joanam/scripts/blob/master/ldPruning.sh), which removes sites with $r^2$>0.2 in windows of 50 Kbp sliding by 10 Kbp. We conducted the PCA using Plink v2.0 with default parameters (Purcell and Chang 2007; Chang et al. 2015). We used a vcf with 3,685,916 high-quality SNPs sites.

We analysed patterns of population structure using the R package *lostruct* (Li and Ralph 2019) 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 (R Core Team 2013) 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 (Alexander et al. 2009). 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 Simon Martin's General

Genomics GitHub repository (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py).

**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 (Danecek et al. 2011), 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 the Simon Martin's General Genomics GitHub repository (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py). 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 (Danecek et al. 2011). We then used the R package windowscanr v. 0.1 from Hugo Tavares' GitHub repository (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 S1). 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 S1, Figure S1). In contrast, *H. sara* and *H. leucadia* had ≤10% missing data per individual, much less than any other species (Figure S1) and

likely explained by close similarity to the reference genome. The mean coverage of all other samples ranged between 10X and 27X (Table S1, Figure S1).

**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. 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 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 (Kozak et al. 2015).

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; Figure 2a and S2-S22), 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; Figure 2a and S2-S22). Similarly, the position of *H. hewitsoni* in relation to *H. sapho* varied in eight chromosomes (Chr5, Chr6, Chr7, Chr14, Chr16, Chr17, Chr20, Chr21; Figure 2a and S2-S22).

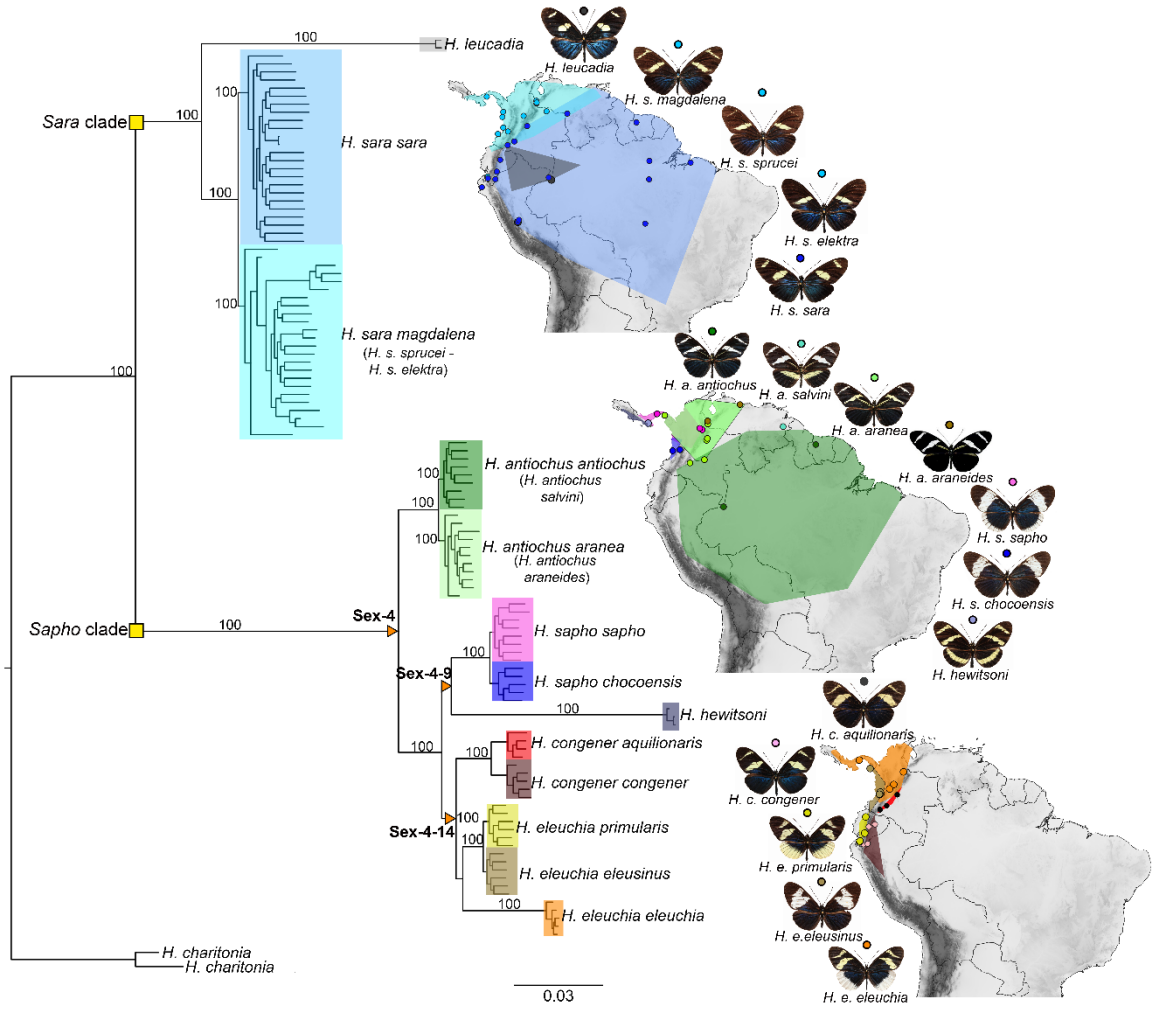


**Figure 1. Phylogeny and distribution of the sara/sapho clade.** Genome-wide ML phylogeny, highlighting the position of the Sara and the Sapho clade, with bootstrap support values on the branches. Orange triangles highlight clades where SA fusions were detected, and the SA fusion is specified. The distribution of all 17 subspecies in the clade from (Rueda-M et al., 2021) 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.

Interestingly, we found sex-specific clustering on Chr4, Chr9 and Chr14 (Figure 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 2b). This sex clustering pattern was also observed on Chr9 for *H. sapho* and *H. hewitsoni* (Figure 2b). In addition, males and females of *H. eleuchia* and *H. congener* also formed a clade in Chr14 (Figure 2b)**.**

**Population structure and shared ancestry**

The PCA grouped samples into three main groups: (i) *H. sara and 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 S23). 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 Andean subspecies of *H. sara* (*H. s. magdalena, H. s. sprucei,* and *H. s. elektra*) and 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 shared ancestry among species (Figure S24).

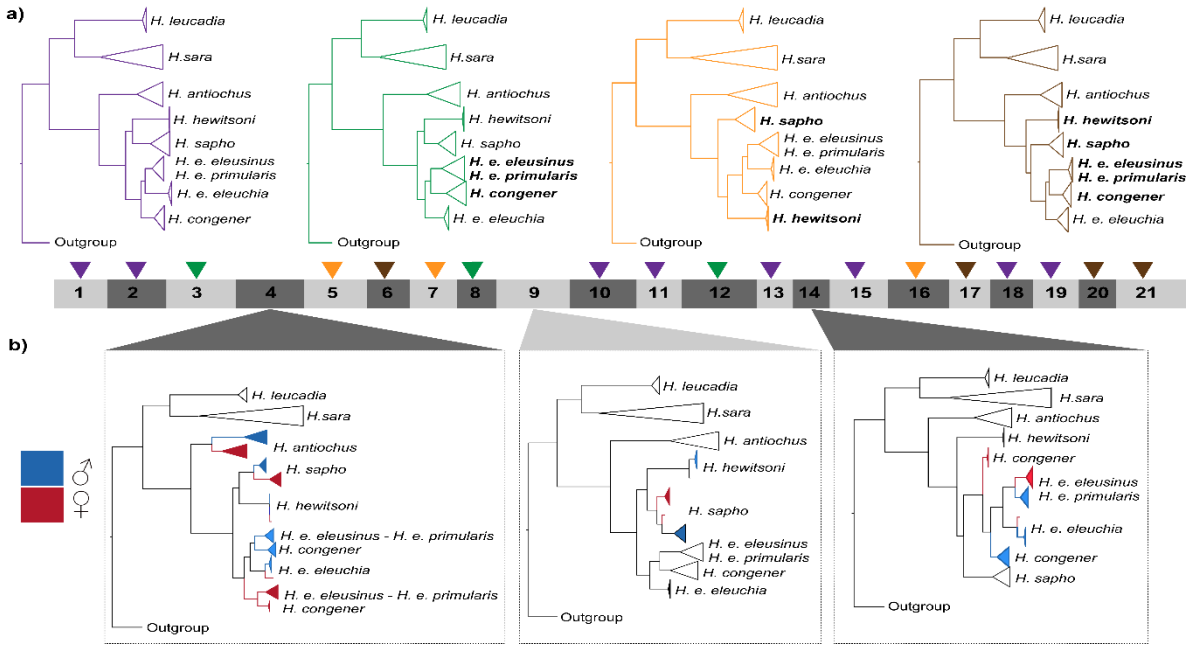


**Figure 2. ML phylogenies inferred genome-wide and per chromosome.** (a) Topologies found across the genome. Purple: genome-wide topology. Green: *H. congener* within *H. eleuchia*. Orange: *H. hewitsoni* as sister to *H. congener* + *H. eleuchia*. Brown: *H. hewitsoni* as sister to *H. congener* + *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 (Figure 3 and Figure S25-S29). 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 (Figure 3a-3b and 3d-3e). In *H. antiochus*, all windows in Chr4 cluster individuals by sex (Figure S27). 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 S28). 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 S29). Strikingly, in all species, sex-grouping windows are scattered along Chr4, Chr9 and Chr14 (Figure 3c, 3f and S28c).

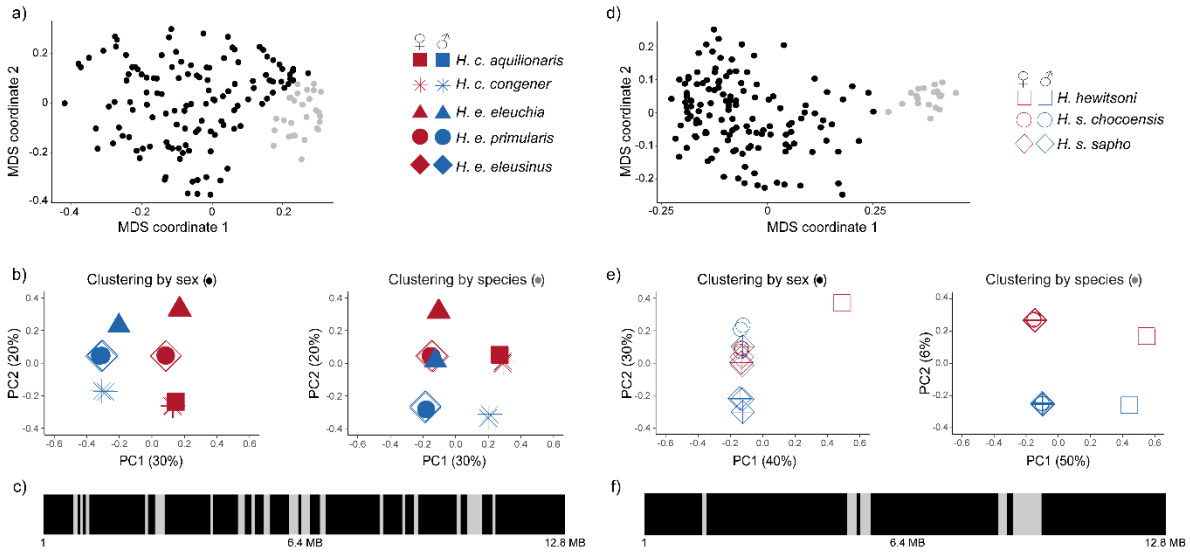


**Figure 3. Local PCA in 100 SNP windows on Chr4 reveals grouping by sex across the entire chromosome. (a)** two-dimensional MDS plot of the distribution of 100 SNP windows along Chr4 for the *H. congener/H. eleuchia* clade, **(b)** PCA plots for the black and grey windows in (a). **(c)** Actual distribution 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.

## 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 4). *H. sara* vs. *H. leucadia* were the least differentiated pair (average $F_{ST}$ = 0.05, Figure 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 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 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 4). Chr14 behaves like Chr9 but only in *H. eleuchia* and *H. congener* (Figure 4). The same pattern was true when we compared $F_{ST}$ and $D_{XY}$ between subspecies (Fig. S30 – S33). 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 S30).

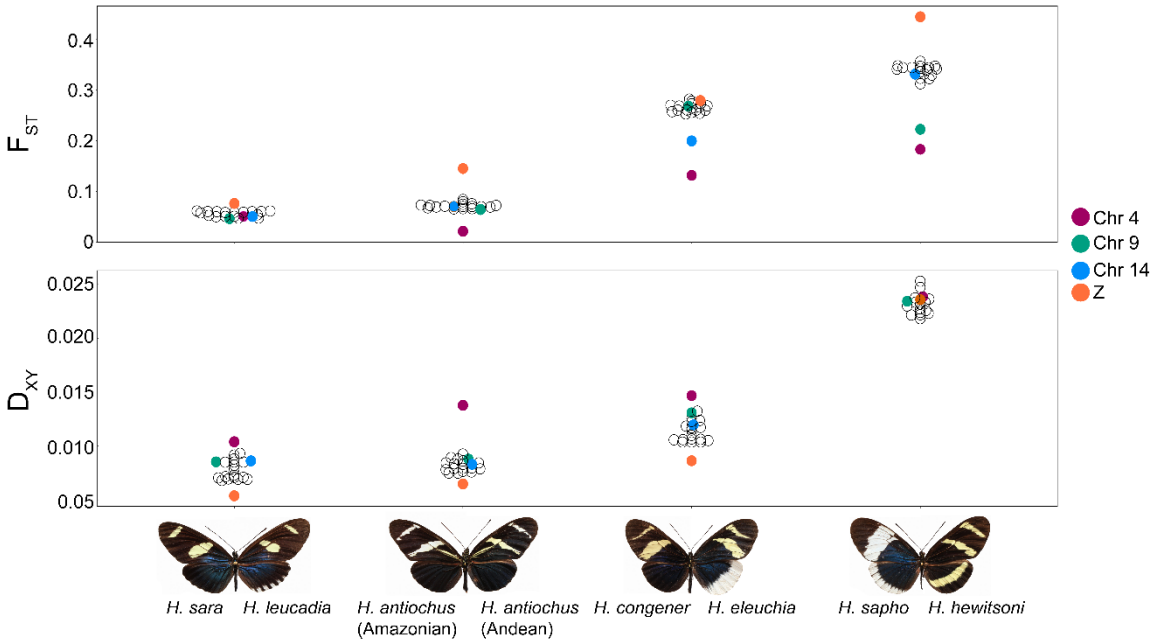

**Figure 4. Genome-wide divergence ($F_{ST}$) between pairs of species in the *sara/sapho* clade.** Each dot represents a chromosome, and chromosomes with evidence of SA fusions are colour coded. Analyses for pairs of subspecies are shown in Figure S30 and S33.

## 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. Because females of Lepidoptera lack crossing over, and their meiosis is achiasmatic, they do not recombine (Suomalainen et al. 1973). This means that, if the Z chromosome is involved in the fusion (Figure 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 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 (Figure 5). Consistently, Chr4, Chr9 and Chr14 showed striking sex-specific differences in heterozygosity in multiple species (Figure 6). 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 6). In *H. eleuchia* and *H. congener*, females also showed very high heterozygosity on Chr14, and in *H. hewitsoni* and *H. sapho,* on Chr9 (Figure 6). Differences between males and females were significant on these 3 chromosomes for all species of the sapho clade (Wilcoxon test, $p < 0.01$, Figures S34-38), 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 S36a). The same was true for Chr14 in *H. eleuchia* and Chr9 in *H. sapho* (Wilcoxon test, $p < 0.01$, Figure S36a).

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 6). These differences were significant in *H. eleuchia*, *H. sapho, H. antiochus* (Wilcoxon test, $p \leq 0.01$; Figure S34). This pattern was not true for Chr14 in *H. eleuchia* and *H. congener,* nor Chr9 for *H. sapho* and *H. hewitsoni* (Figure 6 and S35). 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 S36a). However, this was not true for Chr14 and Chr9 (Figure S36a).

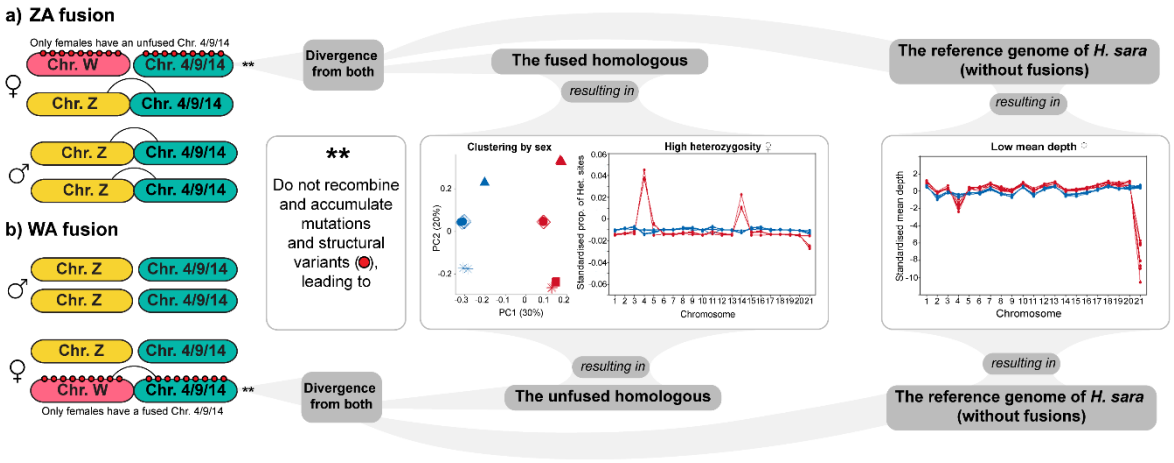


**Figure 5. Scenarios of SA fusions involving either the W or Z chromosome.** Scenario of **(a)** ZA fusion and **(b)** WA fusion for Chr4, Chr9 and Chr14 differentiating the expected pattern by sex. Yellow: Z chromosome.

Pink: W chromosome. Green: autosome. Curved lines: fusions. Red dots: mutations and/or structural variants. Asterisks: chromosomes that do not recombine and may accumulate mutations and/or structural variants thus diverging from their homologous.

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, $p < 0.01$, Figure 7 and S37-S39). 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, $p < 0.01$, Figure 7 and S37-S39). Interestingly, one female of *H. congener* has a region of eight windows on Chr4 with very high sequencing depth (Figure 7). There are also few windows on Chr4, Chr9 and Chr14 where females exhibited both lower heterozygosity and mean depth than males (Figure 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 (Figure 6 and 7). Interestingly, the heterozygosity in females of *H. congener* dropped at the end of Chr14 to values similar to those of the males (Figure S39).

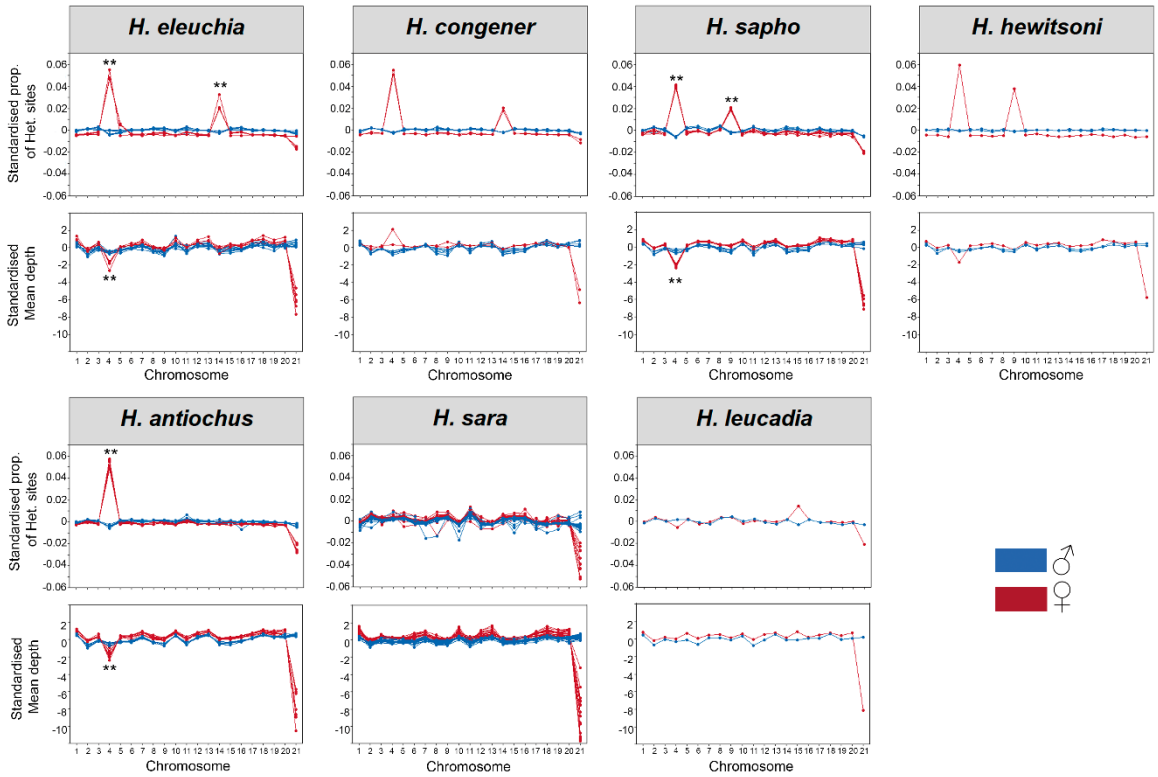


**Figure 6. Patterns of heterozygosity and depth across the genome.** Standardised proportion of heterozygosity sites and standardised mean depth per chromosome in each species (**p<0.01). Each line corresponds to one individual.

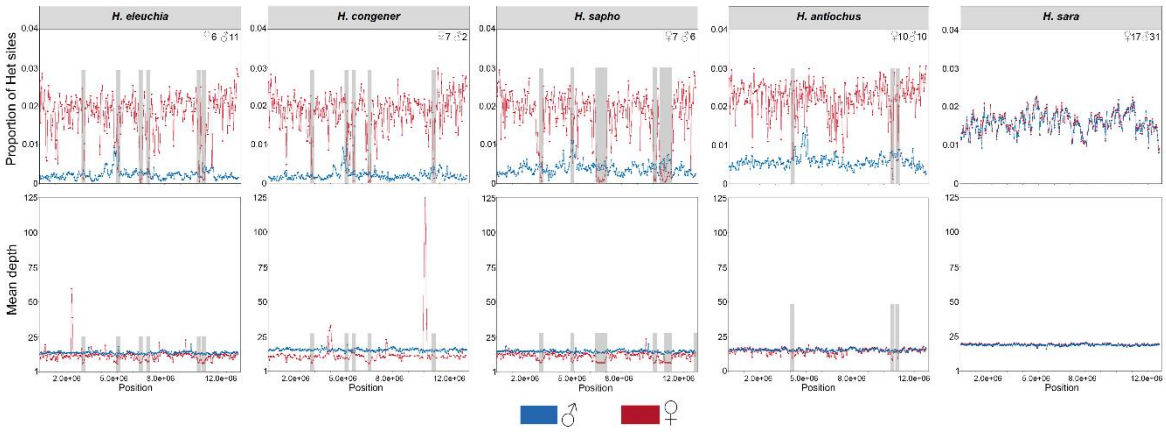


**Figure 7. Patterns of heterozygosity and depth across chromosome 4**. Proportion of heterozygosity sites and mean depth in 50Kb sliding windows in each species. Males are shown in blue and females in red, and their *n* is shown in the top right corner of each panel. Shadowed regions are those where 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, and (ii) high female heterozygosity seen on certain 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. 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 unmappable 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 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 do not recombine anymore (Chr4/9/14 in the ZA fusion and Chr W + Chr.4/9/14 in the WA fusion; Figure 5). Long-read sequencing data will be required to distinguish between these possibilities. No sex chromosome fusions have been reported previously for any *Heliconius* butterflies.

The finding of multiple sex chromosome 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 (Brown et al. 1992). 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 (Jay et al. 2022). 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 (Ross et al. 2009; Pala et al. 2012; Sigeman et al. 2019). 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 (Guerrero and Kirkpatrick 2014).

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 (Kozak et al. 2015), 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 (Beltrán et al. 2007; Kozak et al. 2015) but agrees with a recent whole genome phylogeny based on *de novo* genome assemblies (Cicconardi et al. 2022), 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.

## Data and Resource Availability

The sequence data generated in this study have been deposited in the European Nucleotide Archive database under accession code PRJEB60296 or elsewhere on ENA if obtained from previous studies, as specified in Supplementary Data Table S1 for each individual. Reference genomes for *H. sara* can be found in https://tolqc.cog.sanger.ac.uk/durbin/jiggins/Heliconius_sara/. All records associated to the individuals used for this study are available in the *Heliconius* Earthcape database (https://heliconius.ecdb.io/106).

## Acknowledgments

We also thank “Autoridad Nacional de Licencias Ambientales—ANLA” in Colombia for granting Universidad del Rosario the collecting permit 530 and the Instituto Chico Mendes de Conservação da Biodiversidade in Brazil for granting SISBIO collection licence 59194-1 under which we performed our collecting activities. The HPC Service of Universidad del Rosario (CALDAS) for computing time. This work was funded by Universidad del Rosario (BigGrant IV-FGD005, Fondos Concursables IV-FPD004) and NRM was funded by the Colombian Ministry of Science and Technology (MinCiencias – doctoral scholarship 727).

## Author contributions

NR-M, JM and CS conceived the study. NR-M, CS, CP-D, CJ, OMC, KK, CFA, and JR collected the specimens. SMC, RD, CJ assembled the *H. sara* genome, GM mapped the individuals to the *H. sara* genome reference. NR-M, CS and JM

analyzed the data. NR-M, CS, JM, and CP-D drafted the manuscript. All authors read and approved the final version.

## References


Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.

Beltrán M, Jiggins CD, Brower AVZ, Bermingham E, Mallet J. 2007. 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.* 92:221–239.

Bressa MJ, Papeschi AG, Vítková M, Kubícková S, Fuková I, Pigozzi MI, Marec F. 2009. Sex chromosome evolution in cotton stainers of the genus Dysdercus (Heteroptera: Pyrrhocoridae). *Cytogenet. Genome Res.* 125:292–305.

Broad I. 2019. Picard Toolkit. Available from: https://github.com/broadinstitute/picard

Brown KS. 1981. The biology of Heliconius and related genera. *Annu. Rev. Entomol.* 26:27–56.

Brown KS, Emmel TC, Eliazar PJ, Suomalainen E. 1992. Evolutionary patterns in chromosome numbers in neotropical Lepidoptera. I. Chromosomes of the Heliconiini (family Nymphalidae: subfamily Nymphalinae). *Hereditas* 117:109–125.

Carabajal Paladino LZ, Provazníková I, Berger M, Bass C, Aratchige NS, López SN, Marec F, Nguyen P. 2019. Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea.Barluenga M, editor. *Genome Biol. Evol.* 11:1307–1319.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:s13742-015-0047–0048.

Charlesworth D, Charlesworth B. 1980. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* 35:205–214.

Chow W, Brugger K, Caccamo M, Sealy I, Torrance J, Howe K. 2016. gEVAL — a web-based browser for evaluating genome assemblies. *Bioinformatics* 32:2508–2510.

Cicconardi F, Milanetti E, Castro ÉCP de, Mazo-Vargas A, Belleghem SMV, Ruggieri AA, Rastas P, Hanly J, Evans E, Jiggins CD, et al. 2022. Evolutionary dynamics of genome size and content during the adaptive radiation of Heliconiini butterflies. :2022.08.12.503723. Available from: https://www.biorxiv.org/content/10.1101/2022.08.12.503723v1

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.

Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, Merrill RM, Joron M, Mallet J, Dasmahapatra KK, et al. 2016. Major Improvements to the Heliconius melpomene Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3 GenesGenomesGenetics* 6:695–708.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–498.

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

Flores SV, Evans AL, McAllister BF. 2008. Independent Origins of New Sex-Linked Chromosomes in the melanica and robusta Species Groups of Drosophila. *BMC Evol. Biol.* 8:33.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. Available from: http://arxiv.org/abs/1207.3907

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Comput. Biol.* 15:e1007273.

Guerrero RF, Kirkpatrick M. 2014. Local adaptation and the evolution of chromosome fusions. *Evolution* 68:2747–2756.

Hill J, Rastas P, Hornett E, Neethiraj R, Clark N, Morehouse N, La M, Celorio-Mancera M de la P, Cols J, Dircksen H, et al. 2019. Unprecedented

reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci. Adv.* 5:eaau3648.

Hora KH, Marec F, Roessingh P, Menken SBJ. 2019. Limited intrinsic postzygotic reproductive isolation despite chromosomal rearrangements between closely related sympatric species of small ermine moths (Lepidoptera: Yponomeutidae). *Biol. J. Linn. Soc.* 128:44–58.

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

Jiggins CD, Lamas G. 2016. The Ecology and Evolution of Heliconius Butterflies. Oxford University Press Available from: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199566570.001.0001/acprof-9780199566570

Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, Absher DM, Grimwood J, Schmutz J, Myers RM, et al. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079–1083.

Kozak KM, Wahlberg N, Neild AFE, Dasmahapatra KK, Mallet J, Jiggins CD. 2015. Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic Heliconius Butterflies. *Syst. Biol.* 64:505–524.

Král J, Kořínková T, Krkavcová L, Musilová J, Forman M, Herrera IMÁ, Haddad CR, Vítková M, Henriques S, Vargas JGP, et al. 2013. Evolution of karyotype, sex chromosomes, and meiosis in mygalomorph spiders (Araneae: Mygalomorphae). *Biol. J. Linn. Soc.* 109:377–408.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: http://arxiv.org/abs/1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li H, Ralph P. 2019. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics* 211:289–304.

Lucek K. 2018. Evolutionary Mechanisms of Varying Chromosome Numbers in the Radiation of Erebia Butterflies. *Genes* 9:166.

Lucek K, Augustijnen H, Escudero M. 2022. A holocentric twist to chromosomal speciation? *Trends Ecol. Evol.*:S0169534722000854.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38:4647–4654.

Matsumoto T, Kitano J. 2016. The intricate relationship between sexually antagonistic selection and the evolution of sex chromosome fusions. *J. Theor. Biol.* 404:97–108.

Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. 2018. Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Mol. Biol. Evol.* 35:1489–1506.

Mongue AJ, Nguyen P, Voleníková A, Walters JR. 2017. Neo-sex Chromosomes in the Monarch Butterfly, *Danaus plexippus*. *G3 GenesGenomesGenetics* 7:3281–3294.

Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, Neven LG, Sahara K, Marec F. 2013. Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc. Natl. Acad. Sci. U. S. A.* 110:6931–6936.

Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B. 2012. Evidence of a neo-sex chromosome in birds. *Heredity* 108:264–272.

Pardo-Manuel de Villena F, Sapienza C. 2001. Female meiosis drives karyotypic evolution in mammals. *Genetics* 159:1179–1189.

Pennell MW, Kirkpatrick M, Otto SP, Vamosi JC, Peichel CL, Valenzuela N, Kitano J. 2015. Y Fuse? Sex Chromosome Fusions in Fishes and Reptiles. *PLOS Genet.* 11:e1005237.

Picq S, Lumley L, Šíchová J, Laroche J, Pouliot E, Brunet BMT, Levesque RC, Sperling FAH, Marec F, Cusson M. 2018. 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* 8:2539–2549.

Purcell S, Chang CC. 2007. PLINK: Whole genome data analysis toolset. Available from: www.cog-genomics.org/plink/2.0/

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from: http://www.R-project.org/

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

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17:155–158.

Schmid M, Feichtinger W, Steinlein C, Visbal García R, Fernández Badillo A. 2003. Chromosome banding in Amphibia. XXVIII. Homomorphic XY sex chromosomes and a derived Y-autosome translocation in Eleutherodactylus riveroi (Anura, Leptodactylidae). *Cytogenet. Genome Res.* 101:62–73.

Šíchová J, Ohno M, Dincă V, Watanabe M, Sahara K, Marec F. 2016. 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.* 118:457–471.

Šíchová J, Voleníková A, Dincă V, Nguyen P, Vila R, Sahara K, Marec F. 2015. Dynamic karyotype evolution and unique sex determination systems in Leptidea wood white butterflies. *BMC Evol. Biol.* 15:89.

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

Smith DAS, Gordon IJ, Traut W, Herren J, Collins S, Martins DJ, Saitoti K, Ireri P, ffrench-Constant R. 2016. A neo-W chromosome in a tropical butterfly links colour pattern, male-killing, and speciation. *Proc. R. Soc. B Biol. Sci.* 283:20160821.

Stamatakis A, Hoover P, Rougemont J. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers.Renner S, editor. *Syst. Biol.* 57:758–771.

Suomalainen E, Cook LM, Turner JRG. 1973. Achiasmatic oogenesis in the Heliconiine butterflies. *Hereditas* 74:302–304.

de Vos JM, Augustijnen H, Bätscher L, Lucek K. 2020. Speciation through chromosomal fusion and fission in Lepidoptera. *Phylosophical Trans. B*:11.

White MJD. 1977. Animal Cytology and Evolution. 3rd ed. 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

Yoshido A, Marec F, Sahara K. 2005. 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* 114:193–202.

Yoshido A, Sahara K, Marec F, Matsuda Y. 2011. Step-by-step evolution of neo-sex chromosomes in geographical populations of wild silkmoths, Samia cynthia ssp. *Heredity* 106:614–624.